Title
Optimal Reconstruction of Cosmological Density Fields

Author
Horowitz, Benjamin Aaron

Publication Date
2019

Peer reviewed|Thesis/dissertation

Optimal Reconstruction of Cosmological Density Fields

By

Benjamin A. Horowitz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Uros Seljak, Chair
Professor Aaron Parson
Professor Martin White

Fall 2019

Optimal Reconstruction of Cosmological Density Fields

Abstract

Optimal Reconstruction of Cosmological Density Fields

by

Benjamin A. Horowitz

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Uros Seljak, Chair

A key objective of modern cosmology is to determine the composition and distribution of matter in the universe. While current observations seem to match the standard cosmological model with remarkable precision, there remains tensions between observations as well as mysteries relating to the true nature of dark matter and dark energy. Despite the recent increased availability of cosmological data across a wide redshift, these tensions have remained or been further worsened. With the explosion of astronomical data in the coming decade, it has become increasingly critical to extract the maximum possible amount of information available across all available scales. As the available volume for analysis increases, we are no longer sample variance limited and existing summary statistics (as well as related estimators) need to be re-examined. Fortunately, parallel with the construction of these surveys there is significant development in the computational techniques used to analyze that data. Algorithmic developments over the past decade and expansion of computational resources allow large cosmological simulations to be run with relative simplicity and parallel theoretical developments motivate increased interest in recovering the underlying large scale structure of the universe beyond the power spectra.

The detailed study of this large scale structure has the potential to shed light on various unanswered questions and under-constrained physical models for the dark sector and the nature of gravity. As we reach higher redshifts with statistically significant samples, the large scale structure can serve as a link between local observations and the cosmic microwave background. These surveys rely on a variety of biased probes, including the lensing and distribution of galaxies, imprints of large scale structure in secondary anisotropies of the CMB, and absorption lines in the spectra of high redshift quasars. These observations are complementary; they probe different scales, have different sources of astrophysical and observational uncertainties, have unique degenercies in parameter space, and require their own methods to extract cosmological parameters from.

In this thesis, I discuss a number of new developments in the analysis of these diverse

cosmological datasets. After introductory material, I discuss work re-examining the lensing of the Cosmic Microwave Background by cluster-sized objects and implement techniques for accurate mass estimation. I demonstrate that this analysis is optimal in the low noise, small scale limit. In the second part, I develop a maximum likelihood formalism for linear density fields, applicable for reconstructing underlying signal from a variety of cosmological probes including projected galaxy fields and cosmic shear, showing that effects of anisotropic noise and masking can be mitigated. Finally, I extend this work to nonlinear observables by using a forward modeling approach for Lyman Alpha forest tomography, finding more accurate cosmic web reconstruction verses existing techniques. The unifying theme of all these works is revisiting existing matter density reconstruction techniques with a critical eye and using new statistical and computational techniques to efficiently perform an unbiased, lower variance, estimate. Included is discussion of the possible impacts of these methods to improve constraints of cosmological parameters and/or astrophysical processes.

Contents

List of Figures

List of Tables

Dedication

I would like to dedicate this thesis to my parents, David and Kathy, who have provided a loving and caring environment over the past twenty seven years for which I feel incredibly thankful. I would also like to give a special thanks to my siblings, Josh and Rebecca, whom which have played a pivotal role in shaping who I am today.

Acknowledgments

I would first like to acknowledge my (now) fiancee, Stephanie Ger, with whom I've shared countless adventures over the course of my academic career including during the research presented within this thesis. It is difficult to imagine completing it without her love and support over the past many years, as well as all our excursions into the Sierras and abroad.

I would also like to thank a number of teachers and professors over the years who have played a direct role in my course of study. From high school, I thank Mark Orsene for whom I attribute much of my early growth in the mathematical sciences and interest in astrophysics. From my undergraduate degree, I would like to thank Prof. Charlie Baltay, Prof. Priya Natarajan, Prof, Francesco Iachello, Prof. Igor Frenkel, Prof. David Poland, and Prof. Robert Zinn for providing me a great deal of mentorship and opportunities to struggle with research over the years.

Here at Berkeley, I would like to thank members of my committee, Prof. Uros Seljak, Prof. Martin White, and Prof. Aaron Parsons, for their support and guidance on this thesis. Of equal importance are the various postdoctoral associates and research scientists who have made the research presented in this thesis possible; Simone Ferraro, Blake Sherwin, Zarija Lukic, and among many others. I particularly thank Grigor Aslanyan, Chirag Modi, Nick Hand, and Yu Feng, whose various computational tools provide the backbone for much of this thesis. I am also particularly grateful for the help desk staff at the National Energy Research Scientific Computing Center for helping solving my various technical problems (mostly of my own making).

I also give thanks to my friends over my time in California, particularly my long-time housemates and fellow graduate students John Groh, Thomas Mittigia, and Chris Akers, as well as my new housemates Tom Zick and Abigail Polin. A key source of support and diversion during my time in graduate school has been rock climbing, and I give special thanks to my longtime partners Stephanie Ger, Tabya Sultan, Tal Peretz, Nico Bauxbaum, Turner Anderson, Audun Mathias Ovstebo, and particularly John Groh with who've I've shared a number of planned and unplanned bivouacs.

I am also appreciative of the number of universities and institutes who have hosted me for various conferences and workshops over the years, including the Yukawa Institute for Theoretical Physics in Kyoto, Osaka University, Academia Sinica Institute of Astronomy and Astrophysics in Taipei, Tsinghua University in Beijing, Nordic Institute for Theoretical Physics in Stockholm, International Center for Theoretical Physics in Trieste, the Royal Observatory in Edinburgh, and the Institut Henri Poincare in Paris. The hospitality of my hosts and/or the organizing committees, and in many cases their efforts to overcome significant language barriers are greatly appreciated. I give a special thanks to the faculty at the Kavli IPMU in Japan where I spend many months of this research, particularly Prof. Khee-Gan Lee and Prof. Masahiro Takada.

Additional acknowledgements, including for data access and project funding, are provided in the relevant project's chapters.

Chapter 1

Introduction

> Many scientists confuse "difficult" with "interesting."
>
> Bohdan Paczyski

A common distinction is drawn between *cosmology* and *cosmography*, where cosmology assumes the Einstien Field Equations (or a parametrized variation) hold and focuses on using data from observations to constrain the underlying physical parameters of the universe (via the Friedmann-Robertson-Walker metric solution) while cosmography is a descriptive approach of mapping the universe where positions of the large scale structure are plotted, like an medieval cartographer's map. These "maps" of the underlying matter density can then be matched with models through various analysis to standard/extended cosmological model or, alternatively, used with more pure astrophysical observations to constrain galactic and/or stellar processes. The recent seminal works of cosmology, such as results from the Planck Satellite [3] or the Baryon Oscillation Spectroscopic Survey [9], have focused entirely on the ubiquitous triangle plots showing how various cross-correlations and power spectra analysis result in additional constraints to parameters like Ω_m and σ_8. From this standpoint, this thesis is focused on the latter cosmography task; there is little emphasis (past the introduction) of constraining cosmological parameters, while the techniques could certainly be extended for that task. Instead the focus is on reconstructing the density fields we indirectly observe. Of course to some extent this does require some a-priori assumption of general relatively (for example, to perform gravitational lensing) but it is not the goal.

In this work I will highlight three projects which have tried to map the matter distribution from a different observables. The writing in each is primarily my own, and all figures in the main text were made by me. In addition to these "cosmographical" works there were two papers from my time as a graduate student, ([82] and [81]) which are more "standard" cosmology papers where physical parameters are constrained; these are not included in this thesis but the curious reader is encouraged to look at those as well.

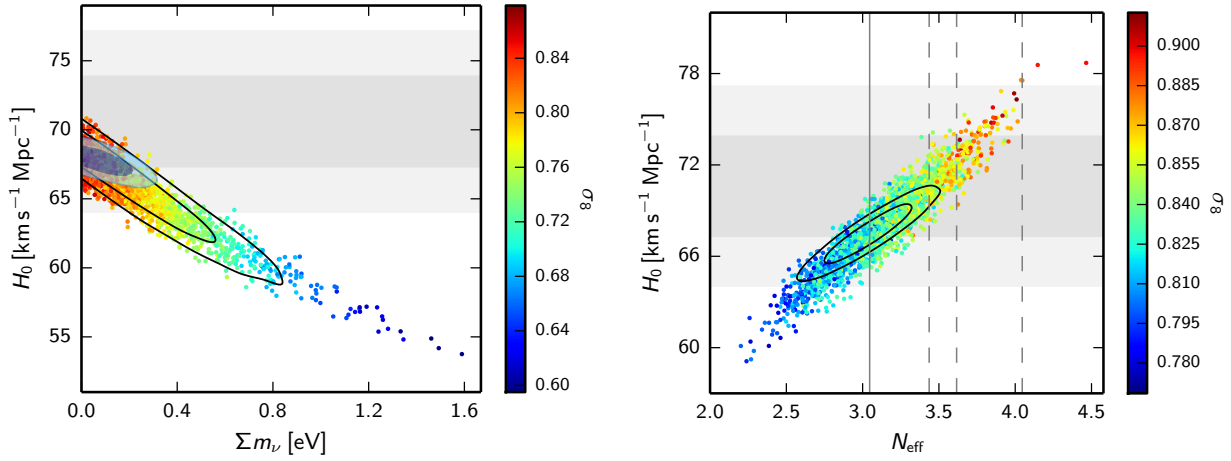


Figure 1.1: A ubiquitous parameter constraint plot from [3], highlighting the power of cosmological to constrain fundamental physics. Dependency of the neutrino mass, $\sum N_\nu$, and number, N_{eff}, on value of the H_0 and σ_8 based on analysis from various cosmological datasets. Grey bands indicates the constraints on H_0 from local observations. Better constraining these parameters is a major focus of modern cosmology, and tensions between CMB and local observations has motivated an increased interest in the field.

1.1 Cosmology Today

Over the past three decades, there has been an explosion in the availability of astronomical data-sets that have changed how we have viewed our place in the universe. What was once an imprecise, rough order of magnitude, field (as described in more detail in Chpt 2.1) has transformed into a "precision" science where parameters are constrained through advanced statistical analysis. In analogy to the particle collider physics of the Large Hadron Collider, surveys are designed in cosmology to optimize the science return. Instead of fine tuning strength of magnetic fields to accelerate particles to probe new energy scales, in cosmology survey strategy/parameters (exposure time, telescope aperture, etc.) are varied to find predicted effects and new physics.

Starting with the discovery of the accelerated expansion of the universe [162, 150] and its potential relationship to quantum fluctuations, there has been an increased interest in using precise cosmology as a constraint in fundamental physics. While certain standard cosmological parameters, like H_0 and σ_8, don't have an immediate connection with the quantum world, others, like the equation of state of dark energy and the properties of neutrinos, do and give insights that aren't possible from lab-based experiments. One particularly notable example of this is from the results of the Planck mission [3] which has provided constraints on the sum of the neutrino masses, Σm_ν, and the number of "effective"[1] neutrinos, $N_{\rm eff}$,

[1]The Standard Model predicts $N_{\rm eff} = 3.046$, where the fractional neutrino is related to the decoupling

which is highlighted in Figure 1.1. This connection between the microscopic and macroscopic is highlighted in the degeneracy present in these relations; better constraining the "pure cosmological" parameters results immediately in better constraints in the Beyond the Standard Model physics.

The other side of this problem of constraining parameters is that within cosmological observations there is tension. Local observations of nearby supernova have found different values of parameters than those inferred from distant observation; this is at odds with a governing idea behind modern cosmology; aptly named the Cosmological Principle. This principle is often summarized as "viewed on a sufficiently large scale, the properties of the universe are the same for all observers."[103] More specifically the principle states the universe on large scales is homogeneous, all local patches should have the same underlying physics and statistical properties within sampling error, and isotropic, the universe has the same physical properties in all directions. Or, perhaps more aptly summarized by Astronomer William Keel as "the universe is knowable and is playing fair with scientists."

This premise can and has been tested by many through examining time and spatial variations of physical "constants", (e.g., [139, 100, 69]) occasionally finding tantalizing clues (e.g., [208]) but none that seem to have survived additional rigor or exceeded random statistical fluctuations. This type of analysis is complementary to a phenomenological form of analysis where exotic physics models, motivated often by some underlying Grand Unified Theory, are implemented into likelihood codes and their parameters constrained. The underlying theory of these analysis is that the laws of physics are still constant throughout the universe, just there is an emergent phenomena from some underlying process we have so far only indirectly observed. Perhaps through implementation of the correct coupling term for dark matter or the correct evolutionary term for dark energy, not only will we discover the true nature of dark matter and dark energy but also explain away apparent tensions in observations. While for some of these there might be an indication of new physics, none have come close commonly accepted definitions of a "detection."[2]

These types of studies are formulaic in that they use well studied cosmological data-sets in similar ways. A model is implemented in a Boltzmann equation solver, meant to model the interactions between particles through specified dynamical couplings and gravity, or other (semi)-analytical model to predict summary statistics such as the power spectra of the CMB or the galaxy correlation function today. This is then compared via a likelihood code to data-sets and parameters estimated.

Of course, summary statistics are limited in two significant ways. First, summary statis-

temperature. Various exotic physics models, such as coupled neutrino or dark radiation theories, predict variations of this value.

[2]The exact threshold for this is fairly subjective and could be defined by different statistical methods. The much hailed, and since retracted, ported discovery of inflationary B-modes in the CMB polarization indicated a 7σ discovery of the scalar to tensor ratio, r, not equaling zero. The detection of the Higgs Boson at the Large Hadron Collider at $\sim$125 GeV was a 6σ detection. The discovery in 2016 of gravitational waves from a binary black hole merger was described as "signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203,000 years."

tics are usually computationally involved to calculate and require careful calculations of associated errors. In the case the large quantities of data found in modern large scale structure surveys it becomes very difficult to scale existing techniques to accurately estimate these quantities. Secondly, since the underlying field has very complex statistical properties it is difficult to create a lossless summary statistic without strong additional assumptions which likely don't hold. For example, the power-spectra is lossless only if the field is purely Gaussian which isn't even the case for the (famously very linear) CMB due to secondary anisotropies from lensing, scattering, and gravitational red-shifting. This motivates a separate component of cosmological analysis; to develop new, more accurate but computationally feasible ways to estimate summary statistics, and to develop new summary statistics which contain additional cosmological information.

Chapter 2

Large Scale Structure

In this chapter we review the key theoretical and observational results from large scale structure, with a particular focus on the cosmological implications of those structures. The goal is to demonstrate the role large scale structure has in connecting the early universe and the current day, which is the motivation for much interest in the field, as well as summarize some of the key statistical tools used in its analysis. By understanding how early fluctuations give rise to the current cosmic web we can take use data from the current day to understand those primordial conditions.

2.1 Observational History

The large scale structure of the universe was first examined by Edwin Hubble shortly after his seminal discovery that many observed "nebula" were extra-galactic [88]. In his work [87] he took a cataloged and classified 400 "extra-galactic nebula" and found they comprised a mostly uniform distribution on the sky. In the following years, a more systematic study of the sky was conducted, expanding the sample to 44,000 "nebula" finding again that the distribution was fairly uniform and conformed to a Gaussian distribution on large scales, with a significant under-density in the "zone of avoidance" along the galactic plane, show in Fig 2.1. However, he noted that on small angular scales, $< 10\,\mathrm{deg}$, there is an excess in galactic number counts above what would be expected from a random Poisson distribution; i.e. galaxies are clumped in certain places.

In the following decades this work was further enhanced by deeper observations over more sky, such as the detection of over a million galaxies using the 0.5m refractor at Lick Observatory in 1967 [176]. A re-reduced version this same data at higher resolution was used to create a vivid map of cosmic structures projected on the sky [169], showing the first view of cosmic web with the now familiar network of filaments and clusters.[1] This data was later used by James Peebles in his seminal work first estimating the two point and three

[1]The authors note that "interested investigators" can get a "copy of the catalog, the raw counts and our corrected counts, on receipt at Princeton of an IBM compatible 2400-ft magnetic tape." [169]

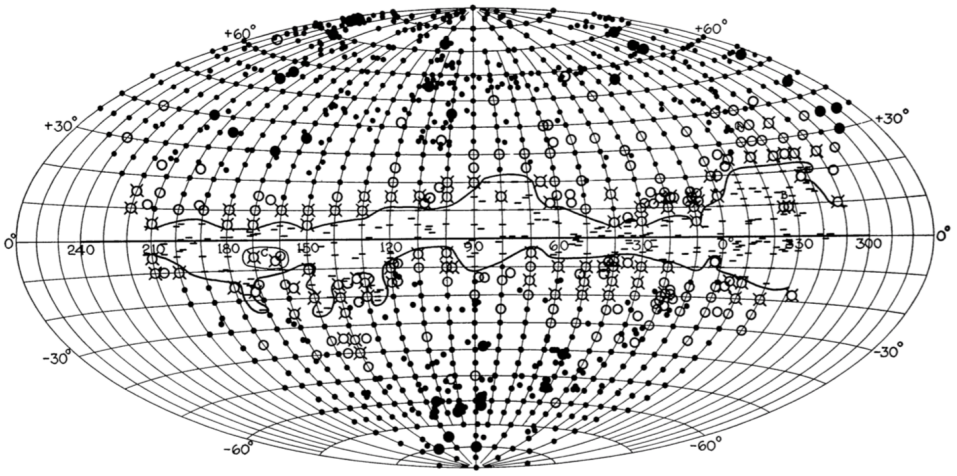


Figure 2.1: An early map (1932) of the extragalactic sky, each mark indicates a pointing/photographic plate from the 6o-inch reflector at Mount Wilson. Larger solid black dots indicated a relative overabundance, and larger empty circles indicate relative underabundance. From [86].

point function of astronomical data [147, 146], which was part of his major contributions leading to his Nobel Prize for Physics in 2019 (coincidentally the target year of this thesis publication). Not only was this work significant as the first use of the power spectrum for cosmological analysis, but it is one of the first uses of "fast digital computers" for calculation of relevant statistics, as opposed to the slower human computers used before this point. From this standpoint these works are arguably the first *modern* cosmological papers; other than the typesetting/notation, it wouldn't look out of place on arXiv today.

Peebles determination of the power spectra (see Section 2.2.1), finding a power-law structure over $O(1\,\mathrm{deg})$ angular separation in the sky, spurred considerable interest in the field. One question of interest was the redshift distribution and correlations present; studies up this point used only the projected fields from photometric surveys which had significant interdependence cosmic structure, magnitude-dependence drop-off, red-shifting out of the filter ranges, etc. The first work to identify cosmic structures in three dimensions and attempt to make full three dimensional maps was in [67], where they took spectroscopic observations of 238 galaxies in the Coma/A1367 Supercluster and immediate environment, shown in Fig 2.3. This was an early example of a complete magnitude-limited spectroscopic survey designed to study cosmic structures, where all objects with a magnitude $m_p < 15$ in the target region were spectroscopically observed and should provide a representative, if biased, sample of the actual matter density. Not only did this data provide a vivid picture of Coma Cluster, A1367, and the inter-cluster structure connecting them, but also the presence of significantly under-dense regions around these structures. An analysis in the Southern Hemisphere [97]

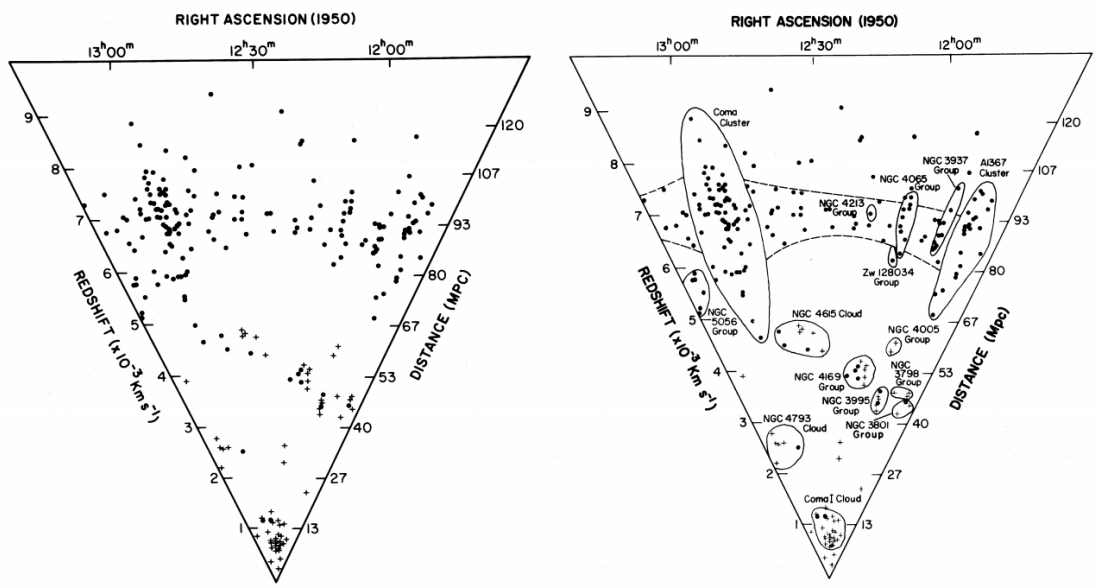


Figure 2.2: A wedge diagram showing the position of galaxies, with redshifts converted to distances assuming $H_0 = 75$ km s^{-1} Mpc^{-1}, projected over 15 deg of declination. The elongation along the line of sight of the cosmic structures was recognized at the time as coming from velocity dispersion within the cluster; now known as "finger of god" effects (see also [90]). From [67].

found similar "holes" where galaxy densities were at least 10 times lower than cluster regions (now known as "voids") as well as "chains of galaxy clusters" (now known as "filaments.")

It is interesting to note that throughout this time, the idea of large scale structure was not commonly accepted. Despite himself making notable catalogues of clusters [218], Fritz Zwicky strongly argued that, while galaxies themselves cluster, the distribution of galaxy clusters was inherently random and uniform with the apparent structure caused by extinction due to intergalactic dust or due to misidentifying of a single cluster as multiple clustered galaxy clusters[217]. Fritz Zwicky's confidence that there are no clusters of galaxy clusters inspired him to explore various possible explanations including massive gravity models. Even into the late 1970's this view was still held by many [11, 59] who attempted to explain away apparent clustering by projection effects and/or extinction from the galactic plane causing differential selection.

A new generation of surveys in the 1980s [105, 42] seemed to have quiet these critics due to the larger data-sets going deeper into the sky. The Harvard Center for Astronomy (CfA) survey [42] spectroscopically observed 2,400 galaxies across both the northern and southern hemispheres, noting that both the "frothy" distribution of galaxies due to the many filamentary structures connecting clusters, as well as the statistical similarity between the northern and southern hemispheres. This paper is notable as one of the early examples

of comparing observations to n-body simulations, in this case a $\Omega_m = 1$ simulation [52], where the connectivity of the cosmic web was compared. Over the next decade, a second CfA survey greatly expanded the footprint of the survey to contain a total of 4391 galaxies, revealing large voids around the Perseus-Pisces Supercluster, as well as a"coherent dense wall" extending at least 170 h^{-1} Mpc [160, 89].

In the new millennium, a new class of multiplexed spectrographs on larger telescopes allowed an explosion of galaxy redshift surveys. The Two Degree Field Galaxy Redshift Survey (2dFGRS, [37]) and Sloan Digital Sky Survey (SDSS, [214]) provided redshifts of over a million galaxies over a large cosmic volume. While other galaxy surveys constrained some cosmological models through qualitative arguments (for example, [51, 50]) it was only with 2dFGRS and SDSS that there was sufficient statistical power (in terms of sky fraction and number density) to make a full statistical analysis worthwhile.[149, 35, 196] The Sloan Survey continues operation in upgraded as of the time of this writing (2019) with currently planned operation into 2020, with power spectrum measurements still informing cosmological constraints [153].

Outside of galaxy surveys, one can also get a unique view of the large scale structure from other probes. In 1971, the UHURU satellite[2] detected x-ray emission not associated directly with any individual galaxy but instead from between galaxies known to be in a cluster [30], resulting eventually in a catalog of such sources. [63, 64] These observations determined that clusters of galaxies are the brightest extra-galactic x-ray sources and can be quite extended, with sources ranging from 200 kpc to 3000 kpc. This was arguably the first direct detection of the intergalactic medium (IGM), which was inferred based on early x-ray observations of the global background [66] as well as by absorption features [68] in the spectra of a high redshift quasar [167] (further discussed in Section 2.3.3). Around this same time, observations of radio sources showed a significant dependence on cluster environment on the observed radio signal which was attributed to a gaseous medium 15 times denser inside clusters compared to the universe's average [44].

While x-ray and radio observation continued to be key in understanding the hydrodynamical properties within clusters throughout the 80s and early 90s, detection of emission from filamentary structures between clusters proved difficult [20]. It was not until the turn of the millennia that filaments were detected across a number of wavelengths, including radio [197], x-ray [166], and microwave [24]. The last of which was inferred based on the apparent inverse Compton Scattering of cosmic microwave background photons off the intergalactic medium, known as the Sunyaev Zeldovich effect [191]. This particular effect has been used as a very powerful probe of large scale structure, both through its power spectra [110] and through well resolved sources [43] (see Sec 2.3.1 for more discussion of cluster cosmology). More recently, the gravitational lensing of the CMB has opened up a new way to study the matter distribution with less dependency on the large scale structure, including via cluster lensing [128, 15] and filament lensing detected in cross correlations [73].

[2]The UHURU mission was one of the earliest space telescopes, launched only three years after the Orbiting Astronomical Observatory, the first such telescope.

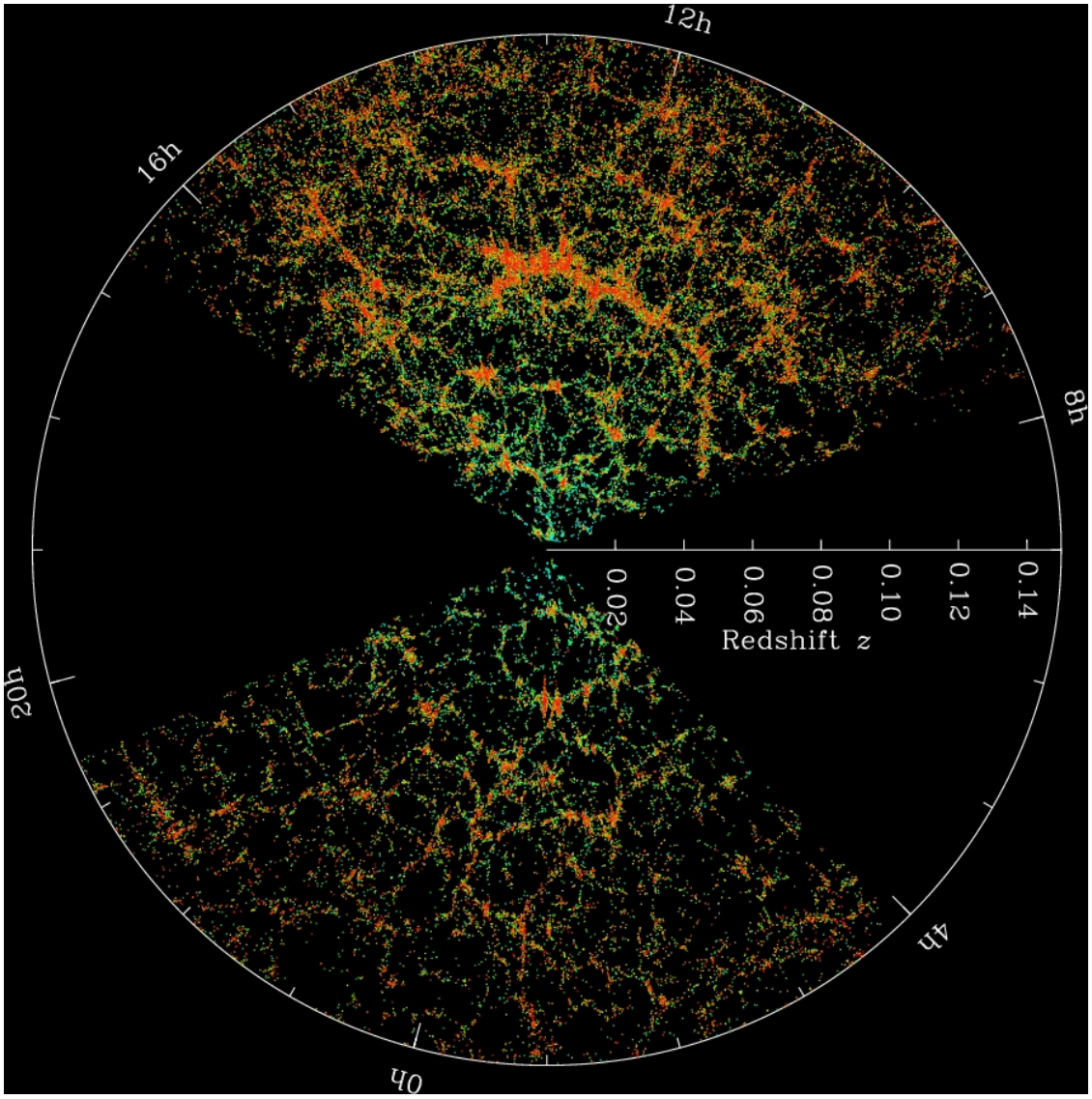


Figure 2.3: The famous pie slices from the Sloan Digital Sky Survey.

2.2 Initial Conditions and Evolution of Large Scale Structure

The origin of large scale structure of the universe across a variety of scales is generally attributed to the gravitational evolution of early primordial fluctuations sourced by an inflaton field. In this section we will briefly summarize the formalism behind the generation of these fluctuations through inflation and evolution of these fluctuations to first order, the Zeldovich approximation [216], leaving higher order perturbation solutions to others (see [16] for a standard reference). Since it is highly doubtful the uninitiated would ever look at this thesis, I will assume a passing familiarity with cosmology and general relativity at the level of a standard undergraduate course.

2.2.1 Matter power spectra

As alluded to in the preceding section, a key focus of large scale structure surveys has been determination of the matter power spectra. The power spectra has become a go to summary statistic to quantify the statistical properties of a density field since, in the Gaussian limit, it should contain all the information about the field relevant for cosmological analysis. While we expect the early universe to be approximately Gaussian, this assumption is known to break down at late time due to the gravitational evolution of the universe. The matter power spectra from various cosmological probes is shown in Figure 5.3. We define the matter power spectra as the two-point function of the matter over-density, $\delta(\boldsymbol{x}) = \rho(\boldsymbol{x})/\bar{\rho}$, i.e.

$$\langle \delta_M(\boldsymbol{x}, t)\delta_M(\boldsymbol{y}, t)\rangle = \frac{1}{V}\int_V d^3z \delta_M(\boldsymbol{x}+\boldsymbol{z}, t)\delta_M(\boldsymbol{y}+\boldsymbol{z}, t). \tag{2.1}$$

We can transform the matter fluctuations into Fourier space as

$$\delta_M(\boldsymbol{x}, t) = \int d^3q \alpha(\boldsymbol{q})\delta_{Mq}(t)e^{i\boldsymbol{q}\cdot\boldsymbol{x}} = F(t)\int d^3q \alpha(\boldsymbol{q})\Delta(q)e^{i\boldsymbol{q}\cdot\boldsymbol{x}} \tag{2.2}$$

where in the last equality we factored out the time dependence from the expression, leaving a term depending on the spectral density per mode.

This assumes we are observing the *real* space positions of the objects though, which is never the case on cosmological distances.

$$\begin{aligned} &\langle \delta_M(r(z)\hat{n}, t(z))\delta_M(r(z')\hat{n'}, t(z'))\rangle \\ &= \int d^3q \delta_{Mq}(t(z))\delta^*_{Mq}(t(z'))\exp i\boldsymbol{q}\cdot(r(z)\hat{n} - r(z')\hat{n'} \\ &= F(t(z))F(t(z'))\int d^3q |\Delta(q)|^2 \exp i\boldsymbol{q}\cdot(r(z)\hat{n} - r(z')\hat{n'} \end{aligned} \tag{2.3}$$

We can re-express this in terms of the power spectra, $P(k)$, and express the low-redshift simplification as

$$\langle \delta_M(r(z)\hat{n}, t(z))\delta_M(r(z')\hat{n'}, t(z'))\rangle = \frac{H_0}{2\pi^2}\int_0^\infty kP(k)\left[\frac{\sin\left((k/H_0)|z\hat{n} - z'\hat{n'}|\right)}{|z\hat{n} - z'\hat{n'}|}\right] \tag{2.4}$$

A related quantity is the variance of the matter fluctuations in a spherical region of radius R,

$$\sigma_R^2(z) = \langle \left(\frac{3}{4^3} \int_{|\boldsymbol{x}|<R} d^3x \delta_M(\boldsymbol{x}, t(z)) \right)^2 \rangle, \tag{2.5}$$

or

$$\sigma_R^2(z) \propto \int_0^\infty P(k) |f(kR)|^2 k^2 dk \tag{2.6}$$

where f is a spherical top-hat function defined by

$$f(kR) = \frac{3}{(kR)^2} \left(\sin kR - kR \cos kR \right). \tag{2.7}$$

The value of this expression for $z = 0$ and $R = 8h^{-1}$ Mpc is known as σ_8 and can be viewed as the overall normalization of this variance, and therefor related to the amplitude of the matter power spectra. The particular value of R was chosen to make this value of roughly order unity (~ 0.8 in the Planck analysis) as well as roughly the cluster scale. Tension in the exact value of this normalization between low redshift observations (such as xray cluster surveys) and that inferred from high redshift CMB observations has driven some interest in beyond the standard model physics between these epochs.

2.2.2 Inflation

Inflation is currently the dominate theory of early universe dynamics which is able to not only source small fluctuations at early times, but also solve a number of significant apparent problems in the standard cosmological model. Perhaps the most perplexing of these is the apparent uniformity of the temperature of the cosmic microwave background. This is difficult to explain if all cosmic expansion is caused by normal matter/radiation as only very small portions of the sky (roughly 2^o) would've been in causal contact at $z \sim 1100$. In addition to this "horizon problem" there is the "monopole problem," where there is an absence of theoretically well motivated magnetic monopoles, and the "flatness problem," where the universe's curvature appears very close to flat. All three apparent problems could be solved if the universe early on undergoes a rapid expansionary phase which would work to dilute any monopoles, flatten any initial curvature, and allow the whole observed universe to be in causal contact at early times before inflation.

Our goal in this section is twofold; (1) we want to study the primary time-vary effect of inflation, the rapid cosmic expansion at early times which "solves" the aforementioned problems, and (2) we want to understand how the presence of this field would give rise to spatial fluctuations. For the former task, our goal will be to relate the presence of a scalar field to the quantities necessary to use the standard Friedmann Equations namely the energy density, ρ and pressure, p, associated with this field. For the latter, we will study how a perturbation to this scalar field will propagate gravitationally.

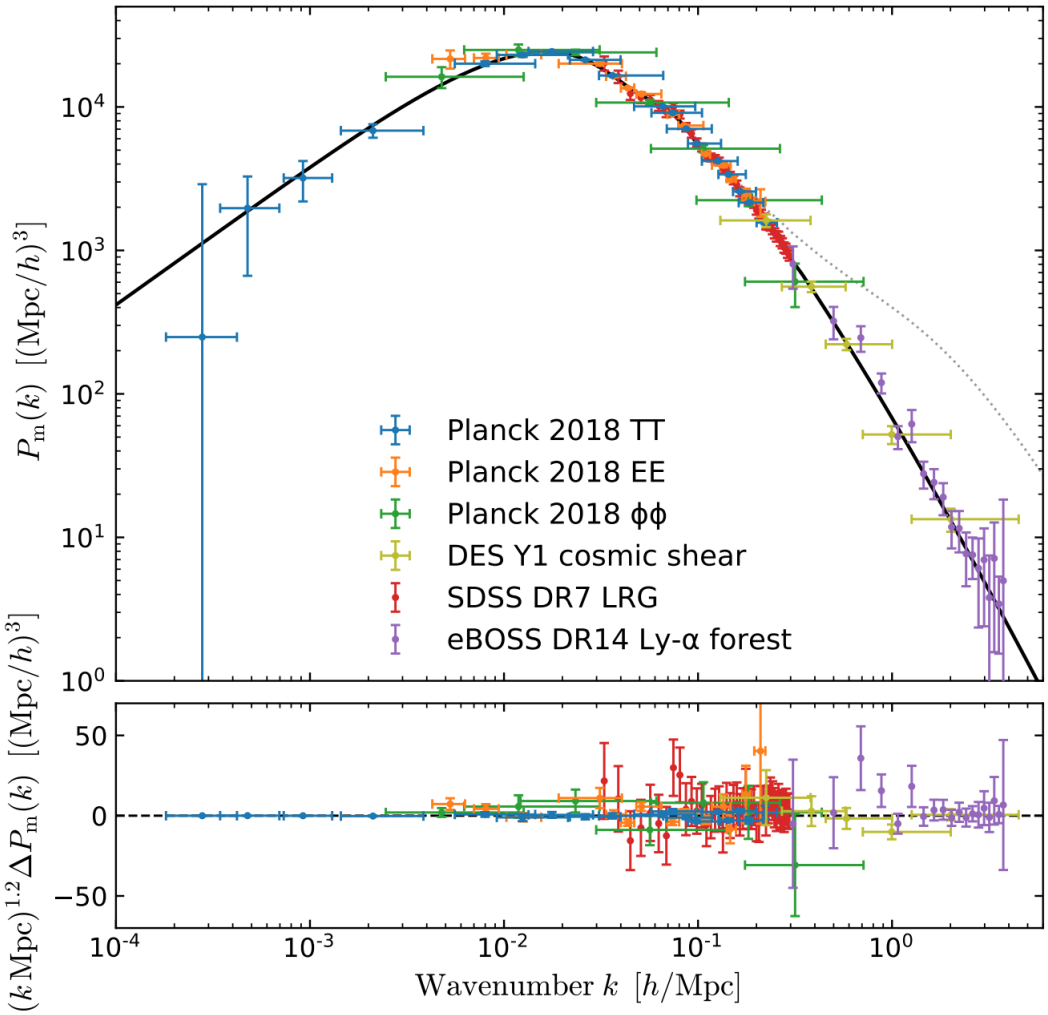


Figure 2.4: Measurements of the matter power spectra at $z = 0$ at a variety of scales and observations. Planck 2018 TT and EE, from [4], relate to the matter power spectra inferred from CMB observations by the Planck Satellite. Planck 2018 $\phi\phi$ is the constraint from the lensing of the CMB by large scale structure discussed in greater detail in Sec 2.3.2.2. DES Y1 cosmic shear, from [201], is a constraint from the distortion of apparent shapes of galaxies caused by lensing from large scale structure, discussed in greater detail in Sec 2.3.2.1. SDSS DR7 LRGs, from [130], are inferred from the distribution of galaxies. eBOSS DR14 Forest, from [32], is a constraint from absorption lines in high redshift quasars and discussed in more detail in Sec. 2.3.3. Figure from [31].

Again assuming the cosmological principle, we theorize a scalar field, the "inflaton" which is primarily only time varying,

$$\phi(\boldsymbol{x}, t) = \bar{\phi}_{\mu\nu}(t) + \delta\phi(\boldsymbol{x}, t), \tag{2.8}$$

where $\delta\phi$ is a small, spatially-dependent, perturbation. The dynamics of a generic scalar field with minimal coupling are governed by the standard action in a curved metric as

$$I_\phi = \int d^4x\sqrt{-g}\left[-\frac{1}{2}g^{\mu\nu}\frac{\partial\phi}{\partial x^\nu}\frac{\partial\phi}{\partial x^\mu} - V(\phi)\right]. \tag{2.9}$$

This can be seen as the sum of a kinetic term and a inflaton potential term, $V(\phi)$. Using the Euler Lagrange equations, i.e. enforcing that the action is stationary, we can find the associated field equations by

$$\frac{\partial V(\phi)}{\partial\phi} = \frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\mu}\left[\sqrt{-g}g^{\mu\nu}\frac{\partial\phi}{\partial x^\nu}\right], \tag{2.10}$$

and associated energy-momentum tensor

$$T^{\mu\nu}_\phi = g^{\mu\rho}g^{\nu\sigma}\frac{\partial\phi}{\partial x^\rho}\frac{\partial\phi}{\partial x^\sigma} - g^{\mu\nu}\left[\frac{1}{2}g^{\rho\sigma}\frac{\partial\phi}{\partial x^\rho}\frac{\partial\phi}{\partial x^\sigma} + V(\phi)\right]. \tag{2.11}$$

We can now extract the pressure and density by noting that

$$\rho = \quad T^{00}_\phi = -\tfrac{1}{2}g^{\mu\nu}\tfrac{\partial\phi}{\partial x^\mu}\tfrac{\partial\phi}{\partial x^\nu} + V(\phi), \tag{2.12}$$

$$\delta_{ij}p = \quad T^{ij}_\phi = -\tfrac{1}{2}g^{\mu\nu}\tfrac{\partial\phi}{\partial x^\mu}\tfrac{\partial\phi}{\partial x^\nu} - V(\phi). \tag{2.13}$$

We can then use the Friedmann equations to show that the first order contribution is

$$H \quad = \tfrac{\dot{a}}{a} = \sqrt{\tfrac{8\pi G\rho}{3}}, \tag{2.14}$$

$$\dot{H} \quad = -4\pi G\dot{\bar{\phi}}^2, \tag{2.15}$$

since we want this field to be rapidly expanding (preferably exponential) at early times for at least $1/H$ we require

$$\left|\frac{\dot{H}}{H}\right|\frac{1}{H} << 1 \tag{2.16}$$

which is equivalent based on Eq. 2.15 to having $\dot{\bar{\phi}}^2 << |V(\phi)|$, i.e. we demand that kinetic term is far less than the potential associated with the field, known as the "slow-roll" condition. To quantify this condition we introduce two variables known as the slowroll parameters,

$$\epsilon = \frac{d(H^{-1})}{dt} = \frac{-\dot{H}}{aH^2}, \tag{2.17}$$

$$\eta = \frac{1}{H}\frac{d^2\bar{\phi}/d^2t}{d\bar{\phi}/dt}. \tag{2.18}$$

We now turn our attention to the effect of the perturbations of the inflaton field on the matter distribution. During inflation the overall energy content of the universe was dominated by this field and so underlying metric perturbations will propagate to the matter field.

To find the metric perturbations induced by the scalar field perturbations (Eq.) we can use how they would propagate to the energy momentum tensor. The conservation equation for the energy-momentum tensor can be written as

$$T^{\mu}_{\nu;\mu} = \frac{\partial T^{\mu}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu}T^{\alpha}_{\nu} - \Gamma^{\alpha}_{\nu\mu}T^{\mu}_{\alpha} = 0. \tag{2.19}$$

To find the equation for how $\delta\phi$ behaves we expand the $\nu = 0$ component to first order and assume a smooth metric. Under these limits the equation simplifies dramatically as

$$\frac{\partial \delta T^0_0}{\partial t} + ik_i\delta T^i_0 + 3H\delta T^0_0 - H\delta T^i_i = 0. \tag{2.20}$$

Note that here we ignored metric perturbations which will act as a sourcing term on the right side of the equation. This approximation should be valid for all sub-horizon modes but will break down for large enough scales; a more detailed analysis is presented in [47]. Rewriting this in terms of $\delta\phi$ in Fourier space we have

$$\ddot{\delta\phi} + 2aH\dot{\delta\phi} + k^2\delta\phi = 0. \tag{2.21}$$

We can work this into a form of a harmonic oscillator by first defining a rescaled field $\tilde{\phi} = a\delta\phi$ which allows the expression to be rewritten as

$$\frac{1}{a}\left[\ddot{\tilde{\phi}} + (k^2 - \frac{\ddot{a}}{a})\tilde{\phi}\right] = 0 \tag{2.22}$$

which has a familiar form to a quantum harmonic oscillator.[3] This allows solution of the form

$$\tilde{\phi}(k,t) = v(k,t)\hat{a}_k + v^*(k,t)\hat{a}^{\dagger}_k \tag{2.23}$$

the variance of these perturbations can be calculated as

$$\tilde{\phi}^{\dagger}(k,t)\tilde{\phi}(k,t)\rangle = (2\pi)^3|v(k,t)|^2\delta^3(\boldsymbol{k} - \boldsymbol{k}') \tag{2.24}$$

where δ^3 is the Dirac delta, expressing that the modes are uncorrelated to first order in the perturbations. Transforming back to $\delta\phi$ we have

[3]The reader here is forgiven for not remembering their introductory quantum mechanics, as I hope I too am forgiven.

$$\phi^\dagger(k,t)\phi(k,t)\rangle = \frac{1}{a^2}(2\pi)^3|v(k,t)|^2\delta^3(\boldsymbol{k}-\boldsymbol{k}') = P_{\delta\phi}(k)(2\pi)^3\delta^3(\boldsymbol{k}-\boldsymbol{k}') \tag{2.25}$$

where $P_{\delta\phi}(k)$ is the power spectrum of the primordial fluctuations. We can solve for $v(k,t)$ using Equation 2.23 and 2.23 to find that

$$v = \frac{e^{-ikt}}{\sqrt{2k}}\left[1-\frac{i}{kt}\right] \tag{2.26}$$

where we used the fact that $\ddot{a} \approx 2a/t^2$. During the rapid expansion of inflation, kt will become very small so we this expression will simplify to

$$v \approx \frac{e^{-ikt}}{\sqrt{2k}}\left[-\frac{i}{kt}\right]. \tag{2.27}$$

Combining this with the relationship in Equation 2.25, we can write the power spectra, $P_{\delta\phi}$, as

$$P_{\delta\phi}(k) = \frac{1}{a^2}\frac{1}{k^3t^2} \approx \frac{H^2}{k^3} \tag{2.28}$$

where we used the approximate relationship $t \approx 1/aH$. This means we expect to have to have a (nearly) "scale invariant" spectra from primordial fluctuations since $k^3P(k)$ is a constant. Since it is deviations from scale invariance we look for in cosmology we can re-write the power spectra as

$$P_{\delta\phi} \propto \frac{k^{n_s-1}}{k^3}\delta_H^2\left(\frac{\Omega_m}{D(a=1)}\right) \tag{2.29}$$

where we have separated out the scale dependence of H into one determined by the growth factor ($D(a=1)$) and an additional scale dependent piece determined by n_s, known as the scalar spectral index. This can be related back to the slow roll parameters as $n_s = 1-4\epsilon-2\eta$.

2.2.3 Evolution of Fluctuations

Some subtlety is needed for understanding the evolution of baryon fluctuations vs. dark matter fluctuations as dark matter decouples from the dominant radiation bath at far earlier time than the baryon distribution. This leads to a significant differential velocity between the baryon and dark matter field at high redshifts which has cosmological implications.[202] In addition, at late times, baryon pressure from collapse and feedback processes will become significant. We will first focus only on the evolution of the dark matter component and assume no back-reaction from the baryons. Studying only dark matter in this approximation is valid enough to derive many cosmological observable to high accuracy.

The evolution of a linear density fluctuation, $\delta = \rho_m/\bar{\rho_m} - 1$, in Eulerian coordinates is determined by the global expansion rate, H, and the (local) force of gravity (via the

continuity and Euler equations) as

$$\frac{d^2\delta}{dt^2} + 2H\frac{d\delta}{dt} - 4\pi G\bar{\rho_m}\delta = 0. \tag{2.30}$$

This mean density however depends on the the expansion rate as well as $4\pi G\bar{\rho}_m = 3\Omega_m(t)H^2/2$. It is convenient to work in units of the scale factor, a, since it can be easily converted to a redshift, so we can re-write this expression as

$$\frac{d^2\delta}{da^2} + a^{-1}\left(\frac{d\ln H}{d\ln a}\right)\frac{d\delta}{da} - \frac{3\Omega_m(a)}{2a^2}\bar{\rho_m}\delta = 0. \tag{2.31}$$

This expression can be interpreted as a damped harmonic oscillator, with the first order term acting like a drag coefficient and the zeroth order term is the instability term. In the linear regime, where $|\delta| << 1$, the evolution of the perturbation is determined by the global expansion history. In particular, for the early universe where dark energy is not yet dominant, the density contrast $\delta(a) \propto a^{3w+1}$ where $w = 1/3$ for the radiation dominated era ($z3500$)[4] and $w = 0$ for the matter dominated era ($z3500$). In general we have a growth rate defined as

$$D_+(a) = \frac{\delta(x,a)}{\delta(x,0)} = \frac{5}{2}\Omega_M \int_0^a \left(\frac{da'}{d\tau}\right)^{-3} da' \tag{2.32}$$

which, for a ΛCDM universe evolved to present day ($a = 1$) is approximately

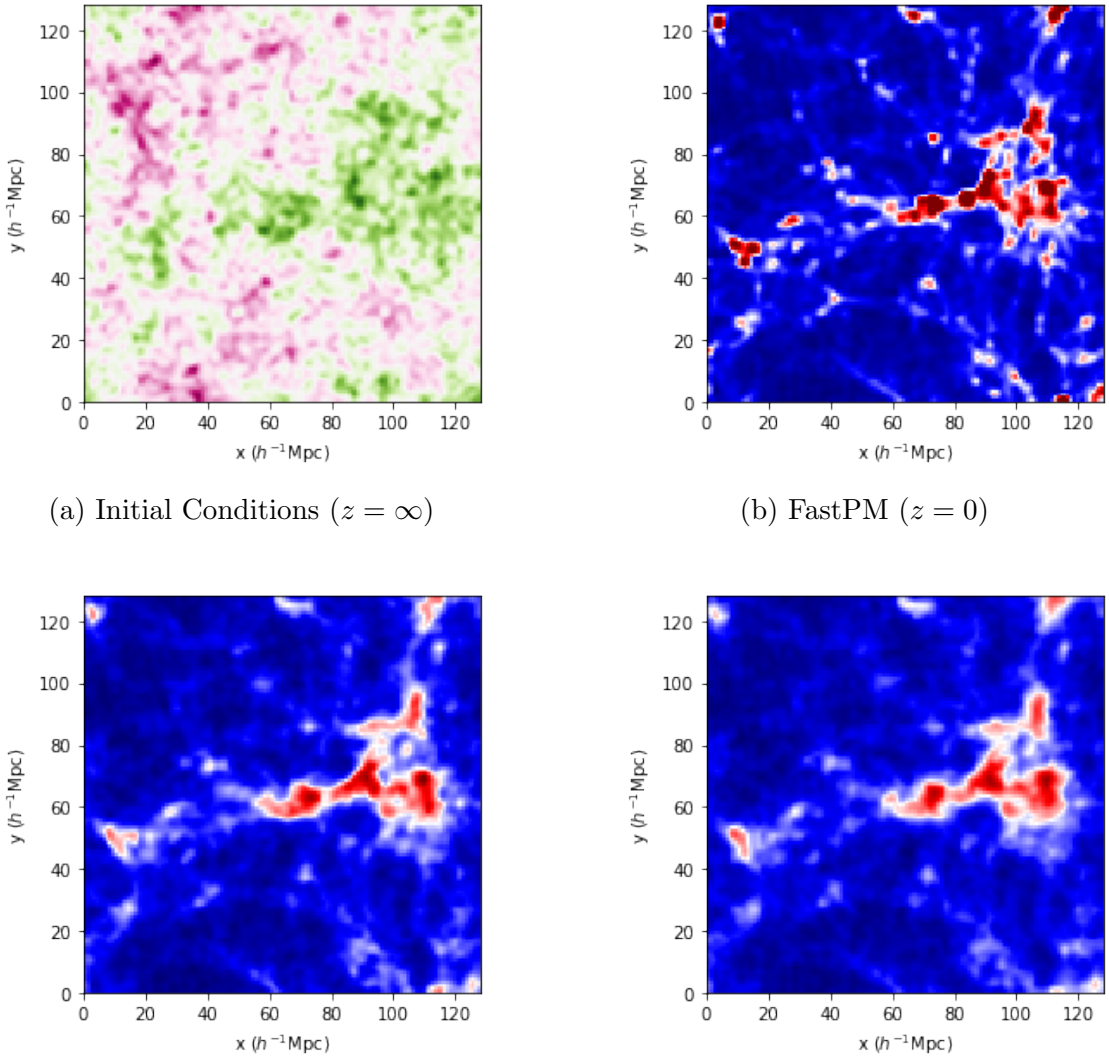
$$D_+(1) \approx \frac{5}{2}\Omega_M \left[\Omega_M^{4/7} - \Omega_\lambda + (1+\frac{1}{2}\Omega_M)(1+\frac{1}{70}\Omega_\Lambda)\right]^{-1}. \tag{2.33}$$

We performed this linear evolution calculation in "Eulerian" coordinates, which are intuitive since they describe the real space distribution of matter. However they are not necessarily optimal when going to higher orders in perturbation theory since correlations between modes due to evolution and due to the initial conditions become inseparable. A popular method is to express things in "Lagrangian" coordinates where position of a particle is described by its initial comoving position, $\boldsymbol{q}$, and a displacement vector, $\boldsymbol{\Psi}(\boldsymbol{q},t)$, i.e.

$$\boldsymbol{x}(\boldsymbol{q},t) = \boldsymbol{q} + \boldsymbol{\Psi}(\boldsymbol{q},t). \tag{2.34}$$

The displacement operator could be expressed exactly in terms of the gradient of the evolving gravitational potential, however this would be numerically difficult to do exactly and instead we will try to recast this into a perturbation problem. We can map between the Eulerian and Lagrangian picture by enforcing mass conservation, i.e. $\rho(\boldsymbol{x},t)d^3x = \rho(\boldsymbol{q})d^3q$, so the Jacobean is

$$J(\boldsymbol{q},t) = \left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}}\right| = \frac{1}{1+\delta(\boldsymbol{x},t)}. \tag{2.35}$$



(a) Initial Conditions ($z = \infty$)

(b) FastPM ($z = 0$)

(c) 1-LPT ($z = 0$)

(d) 2-LPT ($z = 0$)

Figure 2.5: Plots showing projected density fields of the (a) Gaussian initial conditions, and (b) the fully evolved field using a 20 step FastPM. Cosmic structures including voids, filaments, and clusters are clearly visible. Also shown in the figure are the approximate solutions from (c) first order LPT, and (d) second order LPT, both implemented via a one-step FastPM. Notice the changes between first order and second order LPT are qualitatively small.

We can re-write the equation of motion (Eq. 2.30) in Lagrangian coordinates as

$$J(\boldsymbol{q},t)\nabla_x \cdot \left(\frac{\partial^2 \boldsymbol{\Psi}}{\partial t^2} + H\frac{\partial \Psi}{\partial t}\right) = \frac{3}{2}\Omega_m(t)H^2(J(\boldsymbol{q},t)-1), \tag{2.36}$$

where $_x$ denotes the gradient operator with respect to the Eulerian coordinates. Since this a nonlinear equation in Ψ, we use perturbation theory, i.e. Lagrangian Perturbation Theory (LPT), to expand it in first order. It is interesting to note that first order LPT (the Zeldovich Approximation) is generally more accurate than second order Eulerian Perturbation theory in terms of statistical properties of late-time collapsed objects (see, for example, [19]). To linear order we have $J(\boldsymbol{q},t)\nabla_x \approx \nabla_q$ so we can rewrite Eq. 2.36, defining $\psi = \nabla_q \boldsymbol{\Psi}$, as

$$\ddot{\psi} + H\dot{\psi} = \frac{3}{2}\Omega_m(t)H^2\psi, \tag{2.37}$$

which has a similar growth rate solution as the Eulerian linear case (Eq. 2.32). We can go further to second order, writing our solution of the form $\boldsymbol{x}(\boldsymbol{q},t) = \boldsymbol{q} + \boldsymbol{\Psi}^{(1)}(\boldsymbol{q},t) + \boldsymbol{\Psi}^{(2)}(\boldsymbol{q},t)$, where $\boldsymbol{\Psi}^{(1)}$ is the solution to Eq. 2.37 and $\boldsymbol{\Psi}^{(2)}$ described via the shear terms of the first order expression as;

$$\boldsymbol{\nabla}_q \cdot \boldsymbol{\Psi}^{(2)}(\boldsymbol{q},t) = \frac{1}{2}\left(\frac{D_2}{D_+}\right)^2 \sum_{i\neq j}\left[\boldsymbol{\Psi}^{(1)}_{i,j}\boldsymbol{\Psi}^{(1)}_{j,j} - \boldsymbol{\Psi}^{(1)}_{i,j}\boldsymbol{\Psi}^{(1)}_{j,i}\right] \tag{2.38}$$

where D_2 is the second order growth factor which for a ΛCDM universe is approximately

$$D_2(a) \approx -\frac{3}{7}(D_+(a))^2\Omega_M^{-1/143}. \tag{2.39}$$

Lagrangian Perturbation Theory is often the backbone of both analytical calculations and numerical simulations in cosmology. The FastPM framework [58], for example, utilizes LPT to initially evolve the particles till nonlinear effects require a more complete simulation. In Figure 2.5, we show the initial density field and the evolved field with a variety of possible schemes. Through all approximation schemes the cosmic web is clearly visible. The classification of cosmic structure within this Zeldovich approximation is discussed and used in Chapter 5.

2.2.4 Halo Model

In order to map the theoretical understanding of the matter power-spectrum to actual observations we need to account for the bias in the tracer field. For a galaxy survey, linear bias (i.e. $P_{gg}(k) = b_1 P_{mm}(k)$) will work for large scales but at smaller scales galaxies cluster preferentially in halo environments. These halos form through spherical collapse of dark matter perturbations, where dark matter particles will separate from the global expansion (the Hubble flow) to form a "separate universe." The dynamics of the initial collapse can

be understood by modeling the region as an isotropic/homogeneous universe with a higher value of Ω_M, up until the halo reaches an equilibrium due to the viral theorem.

The actual ensemble statistics of these collapsed structures depend on nonlinear gravitational interactions and requires numerical simulations to calculate in detail. However, a fairly accurate analytical simplification, the Press-Schechter formalism, has become a standard tool in understanding these statistics analytically and works surprising well over a range of masses. In this formalism we model the overdensity as a closed universe with matter density Ω_m, with associated Friedmann equation

$$\frac{1}{a}\frac{da}{dt} = H_0 \left(\Omega_m a^{-3} + (1-\Omega_m)a^{-2}\right)^{1/2}. \tag{2.40}$$

Solving the collapse equation to second order one can find that the scale factor evolves as

$$a = \left(\frac{3}{4}\right)^{2/3} \left(\frac{3a_i}{5\rho_i}\right) (\theta - \sin\theta)^{2/3}, \tag{2.41}$$

where θ is a rescaled time parameter set such that $\theta = 2\pi$ corresponds to the the structure fully collapsed. So we have the scale of collapse as

$$a_{col} = \left(\frac{3}{4}\right)^{2/3} \left(\frac{3a_i}{5\rho_i}\right) (2\pi)^{2/3} \approx 1.686\frac{a_i}{\delta_i}, \tag{2.42}$$

so any region with initial density fluctuation greater than or equal to $\delta_c = 1.686$ will collapse. With this threshold we can calculate the statistics of collapsed objected given $\sigma(R)$ (or equivalently $P(k)$, see Eq. 2.5). Since observed properties depend primarily on the mass of the halo, we want to determine number of fluctuations of scale $M = 4\pi R^3/3$ and express the fluctuations on this mass scale, i.e. $\sigma_M = \sigma_{4\pi R^3/3}$. We can calculate the fraction of mass in halos of size $> M$ as

$$N(M) = \frac{1}{\sqrt{2\pi}\sigma(M)} \int_{\delta_c}^{\infty} d\delta \exp\left(-\frac{\delta}{2\sigma^2(M)}\right) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\nu}{\sqrt{2}}\right)\right], \tag{2.43}$$

where erf is the errror function and $\nu = \delta_c/\sigma(M)$. Notice though as we take $M \to 0$ the fraction of matter in collapsed structures goes to $\frac{1}{2}$, which is nonphysical as by late time all matter should be in collapsed structures. To avoid this problem we ad hoc multiply this expression by 2.

We can now use this expression to relate to observations of clusters to constrain the cosmological model. The target quantity we want is the number of halos per mass bin dM, so we calculate the differential number density of halos as

$$\frac{dn}{d\ln M} = \frac{\rho_m}{M}\frac{d}{d\ln M}\left[1 - \mathrm{erf}\left(\frac{\nu}{\sqrt{2}}\right)\right] = \sqrt{\frac{2}{\pi}}\frac{\rho_m}{M}\frac{d\ln\sigma^{-1}}{d\ln M}\nu\exp(-\nu^2/2). \tag{2.44}$$

This formalism is imprecise since a region that is underdense on a given scale may be overdensed when smoothed on a larger scale. While there are some analytical approaches to this problem, most notably excursion set theory where the halo mass function is viewed as a Markov random walk in smoothing length, the most common approach to analyze observations is to use the Press-Schetcher model to motivate fitting functional forms to calibrate observed halos to simulations.

In addition to clustering of galaxies into halos, there is also observed clustering of halos (the "2-halo" term). This effect can be understood through the peak-background split formalism where the large scale density field acts as a "background" fluctuation which enhances the likelihood of a given fluctuation reaching the critical density to form a "peak". This split can be seen as essentially lowering the threshold for collapse where there is a local maxima of the background fluctuation and raising it when there is a local minima. In the case of the Press-Schechter formalism this bias can be found analytically as a function of mass as $b(M) = 1 + \frac{\nu^2 - 1}{\delta_c}$.

2.3 Tracers of Large Scale Structure

As discussed in Section 2.1, there are numerous tracers of large scale structure including galaxy fields, x-ray emission, radio emission, tSZ, gravitational lensing, flux, etc. In this section we focus on the tracers discussed later in the main body of the thesis; clusters, gravitational lensing signal, and Lyman Alpha flux.

2.3.1 Clusters

In Sec 2.2.4, we discussed the Press-Schetcher halo mass function and showed its sensitivity to cosmology via its dependence on the matter power spectra. In general the halo mass function is a very powerful prediction of the Λ-CDM model and is sensitive to the overall matter density, Ω_m, sum of neutrino masses, ΣM_ν, and the amplitude of density fluctuations, σ_8.[163] However, in order to calculate this distribution from observations we need to ascertain their mass which is difficult to estimate. While in certain cases the geometry lines up well for strong lensing signatures to be used to infer their mass, we are usually left to study different proxies of mass, such as X-ray luminosity or Sunyaev-Zeldovich signal, and correlate them with gravitational effects of the cluster, such as galaxy-cluster lensing. The uncertainty in the calibration of these relations now limits the constraining power of cluster counts as a cosmological probe [131].

Information from clusters have been shown to be a key element in breaking degenerecies present in Cosmic Microwave Background power spectra data alone by providing an independent probe of the structure of formation. One of the notable examples is the degeneracy between the amplitude of the primordial matter power spectra, A_s, the optical depth to the surface of last scattering, τ, and the mass of neutrino species. While measurement of the low-

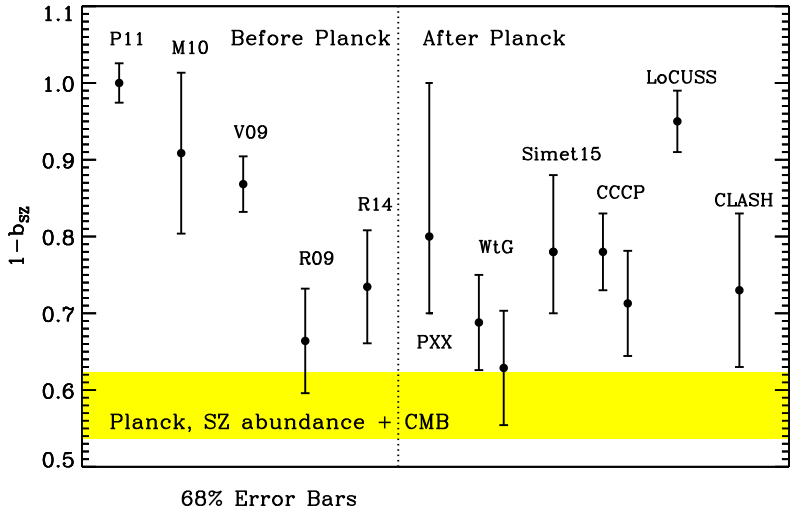


Figure 2.6: Calibration of hydrostatic bias using various techniques and datasets. The wide spread in values corresponds to similarly varying constraints on σ_8 and Ω_m. From [148] .

l polarization power spectra helps to reduce some of these degenerecies, these observations will eventually be fundamentally limited by cosmic variance. [3]

One of the key assumptions in the study of clusters via their tSZ or xray signal is that the cluster halo is spherically symmetric and in hydrostatic equilibrium [165]. This assumption can be understood via the Euler equation

$$\frac{dv}{dt} = -\nabla\Phi - \frac{1}{\rho}\nabla P, \tag{2.45}$$

where Φ is the cluster gravitational potential, P is the (gas) pressure, and ρ is the density. Hydrostatic equilibrium occurs when $\frac{dv}{dt} = 0$; i.e. the three dimensional acceleration of the gas, resulting from the sum of hydrodynamical and gravitational forces, is zero. If one further assumes that the cluster is radially symmetric, with temperature profile $T(r)$ and gas density profile $\rho(r)$, the mass inside a given radius can be calculated as

$$M_{HSE}(r) = -\frac{k_B T(r) r}{\mu G m_p}\left[\frac{d\log\rho(r)}{d\log r} - \frac{d\log T(r)}{d\log r}\right], \tag{2.46}$$

where μ is the mean molecular weight. In observations of clusters via xray, the xray luminosity is a direct tracer of the temperature of the gas, while in tSZ measurements one is probing the pressure profile (which can be related back to the temperature profile). However, recent observations [161] and simulations [113] have shown that most halos are not in HSE and that bulk motion of gas contribute up to 20% of the pressure within clusters. Such

gas motions are thought to be driven by a combination of continuing accretion of gas onto clusters along filaments as well as mergers and supersonic motions of galaxies through the ICM [113]. Shocks caused by this motion can leave behind large wakes that are comparable in size to the length of the cluster [104].

All these effects can contribute to hydrostatic bias within clusters which can be measured by comparing x-ray/tSZ luminosity with lensing signal. A common approach to mitigate this effect is to calibrate out this offset via introduction of a new parameter b_{SZ} (or b_{xray}) where $M_{SZ} = (1 - b_{SZ})M_{true}$. This offset can be calibrated through using a select group of clusters with accurate lensing signal in which to infer the true mass (such as in [123]). However, depending on the methods used to extract this mass and the subset of cluster lenses selected, this calibration can vary significantly. Current cosmological constraints using clusters have been dominated by error on the calibration [148], as shown in Figure 2.6.

2.3.2 Gravitational Lensing

Gravitational lensing, the distortion of the apparent images of distant sources due to intervening matter, has become in recent years a key component of cosmological analysis. It is often grouped into two "types"; strong lensing, qualitatively defined as when the distortion is visible to the eye (either due to extreme arc-like geometry and/or multiple images), and weak lensing which is subtle and usually only identifiable from a statistical ensemble of many sources. In this work we focus on the weak form as it is the type often studied in large volume surveys, both galaxy imaging and cosmic microwave background. Weak gravitational lensing has the advantage of directly probing the underlying gravitational field over large volumes, as opposed to galaxy surveys, cluster counts, and Lyman alpha forest, where the cosmological constraints have high degeneracy with bias models. Weak gravitational lensing is not without its own biases and other systematics (shear bias, magnification bias, intrinsic alignments) which depend on baryon physics but these generally enter beyond leading order in terms of the cosmology constraints.

2.3.2.1 Galaxy Shear

The first order effects of weak lensing are convergence (magnification) and shearing (change of ellipticity) of images.[4] We can describe both effects as a shear matrix

$$M = \begin{pmatrix} \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & \kappa - \gamma_1 \end{pmatrix}, \tag{2.47}$$

where κ is the convergence and γ_i are the shear fields (one for each direction on the sky). This matrix has a direct relation to the effect of lensing on the quadruple moment of

[4]The next order effects are known as "flexion" and involve the change in the octopole moment within the observed image.

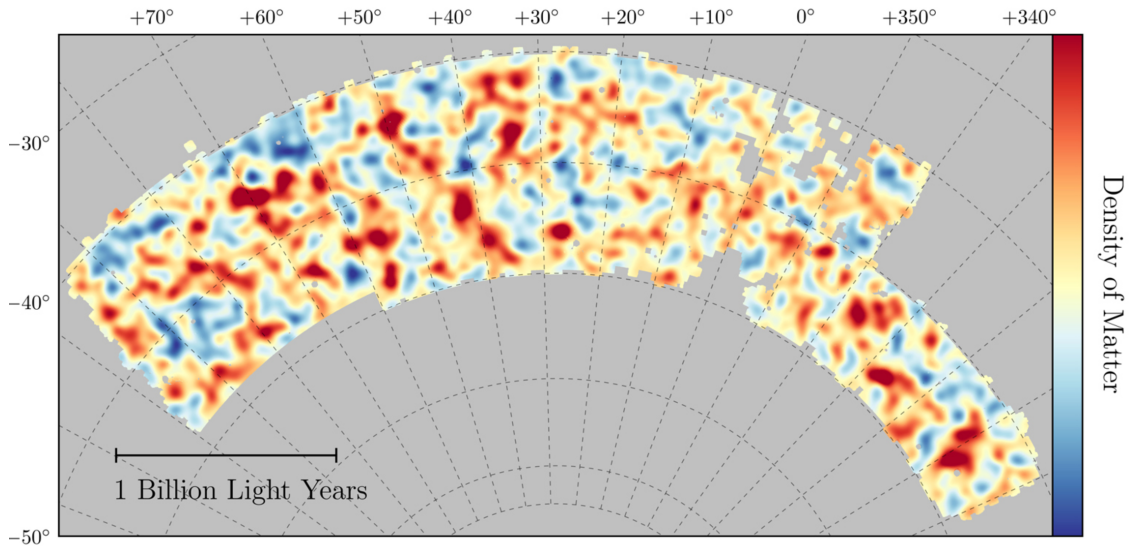


Figure 2.7: Convergence map from the Dark Energy Survey, a cosmic shear survey. The map covers roughly 1/30th of the entire sky and is proportional to the nearby projected mass distribution.

a galaxy's luminosity, Q, with the apparent (lensed) quadrople moment, $\tilde{Q}$, given as

$$\tilde{Q}_{ab} = Q_{ab} + M_{ac}Q_{cb} + M_{db}Q_{ad}. \tag{2.48}$$

In the case of a source at distance r_S in direction $\hat{n}$, with L intervening point sources located at $\boldsymbol{x}_L$ at distance r_L with mass M_L, we can write the shear matrix $M_{ab}(r_S, \hat{n})$ as a sum,

$$M_{ab}(r_S, \hat{n}) = \sum_L \frac{4M_L G r_L(r_L - r_S)}{r_S a_L} \left[\frac{\delta_{ab}}{|\boldsymbol{y}_L|^2} - \frac{2y_{La}y_{Lb}}{|\boldsymbol{y}_L|^4} \right], \tag{2.49}$$

where $\boldsymbol{y}_L = \boldsymbol{x}_L - \hat{n}(\hat{n} \cdot \boldsymbol{x}_L)$ and a_L is the scale factor at the time the light from the source passes the lens. Direct application of this formula is of little use for cosmological analysis as structure is diffuse and extended; rewriting as an integral over the line of sight in terms of a lensing potential, ψ, we have

$$M_{ab}(r_S, \hat{n}) = 2 \int_0^{r_S} \frac{r(r - r_S)}{r_S} \left[\frac{\partial^2}{\partial y_a \partial y_b} \psi(r\hat{n} + \boldsymbol{y}, t) \right]_{\boldsymbol{y}=0, t=t_r} dr. \tag{2.50}$$

Note though for a photometric shear surveys we will not have a well defined position r_S in which to perform this integration, and will instead have some possible probability function $\mathcal{N}(r_S)$ for these galaxies, where the nature of $\mathcal{N}$ will be determined by a combination of selection function and/or photometric redshift. We can now integrate over this function to determine the shear,

$$M_{ab}(r_S,\hat{n}) = 2\int_0^\infty \left[\frac{\partial^2}{\partial y_a \partial y_b}\psi(r\hat{n}+\boldsymbol{y},t)\right]_{\boldsymbol{y}=0,t=t_r} \int_r^\infty \frac{r(r-r_S)}{r_S}\mathcal{N}(r_S)dr_S dr \quad (2.51)$$

$$= 2\int_0^\infty \left[\frac{\partial^2}{\partial y_a \partial y_b}\psi(r\hat{n}+\boldsymbol{y},t)\right]_{\boldsymbol{y}=0,t=t_r} g(r)dr. \quad (2.52)$$

We now want to relate this formula to summary statistics on the sky, in particular the angular auto-power spectra of the shear, $C_{\gamma\gamma}$, and convergence, $C_{\kappa\kappa}$, and thereby have a probe of the lensing potential ψ. We first decompose this potential into plane waves as

$$\psi(\boldsymbol{x},t) = \int d^3q \alpha(\boldsymbol{q})\psi_q(t)e^{i\boldsymbol{q}\cdot\boldsymbol{x}}, \quad (2.53)$$

where $\alpha(q)$ gives the phase information of the density fluctuation. We can now decompose the shear matrix into a shear and convergence component in terms of the plane waves decomposed on the sphere[5],

$$\kappa(\hat{n}) = \sum_{\ell m} a_{\kappa,\ell m} Y_\ell^M(\hat{n}), \quad (2.54)$$

$$\gamma_1(\hat{n}) + i\gamma_2(\hat{n}) = \sum_{\ell m} a_{\gamma,\ell m} Y_\ell^M(\hat{n}), \quad (2.55)$$

with associated coefficients

$$a_{\kappa,\ell m} = 4\pi i^\ell \sqrt{\tfrac{(\ell+2)!}{(\ell-2)!}} \int d^3q q^2 \alpha(\boldsymbol{q}) Y_\ell^{m*}(\hat{q}) \int_0^\infty g(r)\delta\phi_q(t_r)\left(j_\ell(qr) + j_\ell''(qr)\right)dr, \quad (2.56)$$

$$a_{\gamma,\ell m} = -4\pi i^\ell \int d^3q q^2 \alpha(\boldsymbol{q}) Y_\ell^{m*}(\hat{q}) \int_0^\infty g(r)\delta\phi_q(t_r) j_\ell(qr) r^{-2} dr.$$

We want to get create a summary statistic that does not depend on the matter phase information encoded in α, so we can use the orthogonality relationship between components and take the average as

$$\langle a_{X,\ell m} a^*_{X,\ell' m'}\rangle = \delta_{\ell\ell'}\delta_{mm'}C_{XX,\ell} \quad (2.57)$$

for $X \in (\kappa,\gamma)$, where $C_{XX,\ell}$ is the angular power spectra, which is the main observable of these surveys and expressed as

$$C_{\kappa\kappa,\ell} = 4\pi^2 \int_0^\infty q^6 dq \left[\int_0^\infty \delta\phi_q(t_r) j_\ell(qr)(j_\ell(qr) + j_\ell''(qr) g(r)^2 dr\right], \quad (2.58)$$

$$C_{\gamma\gamma,\ell} = \tfrac{4\pi^2(\ell+2)!}{\ell-2} \int_0^\infty q^2 dq \left[\int_0^\infty \tfrac{\delta\phi_q(t_r) j_\ell(qr)}{r^2} g(r)\right]^2. \quad (2.59)$$

In the $\ell >> 1$ limit, we can further simplify this integral by using the properties of the spherical Bessel functions and approximating the integral over r as that over ℓ/q, we get

[5]We make liberal use of the formula $e^{ir\boldsymbol{q}\cdot\hat{n}} = \sum_{\ell m} i^\ell j_\ell(qr) Y_\ell^m(\hat{n}) Y_\ell^{m*}(\hat{q})$, where j_ℓ are spherical bessel functions and Y_ℓ^m are spherical harmonics.

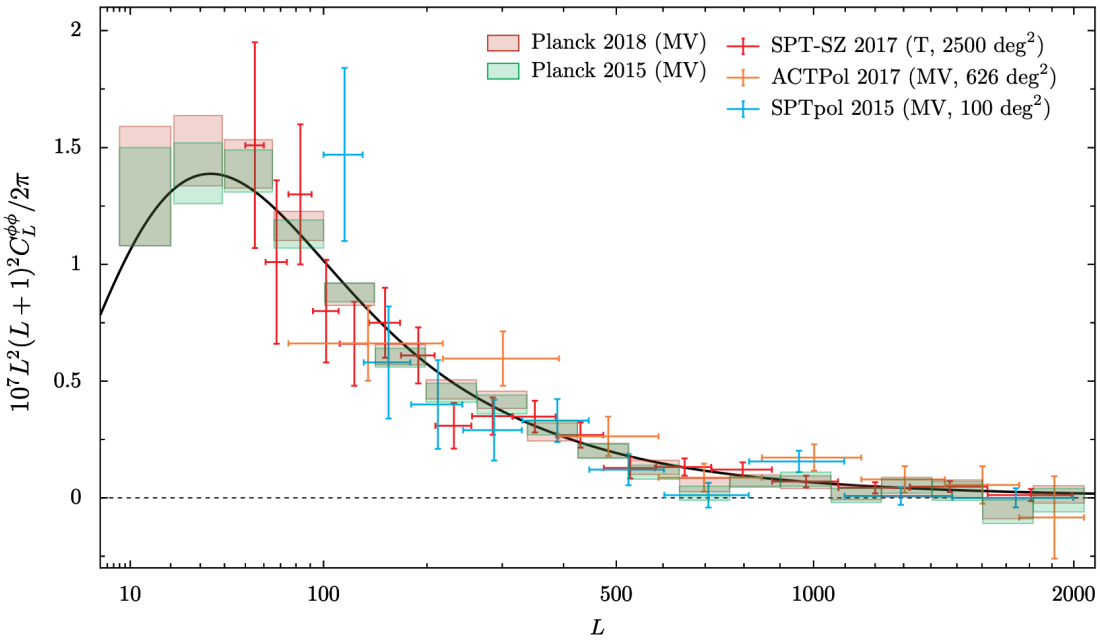


Figure 2.8: Measurements of the CMB lensing potential power spectrum from a number of recent surveys, including Planck 2018 [4], ACTpol [177], SPTpol [187], and SPT-SZ [178]. From [4].

$$C_{\gamma\gamma,\ell} \sim C_{\kappa\kappa,\ell} = 2\pi^3 \ell^4 \int_0^\infty dq q^4 |\delta\phi_q(t_q)|^2 g^2(\ell/q). \tag{2.60}$$

The perturbations $\delta\phi_q$ can be re-expressed in terms of the matter power spectra, see Sec. 2.2.1, to get a more compact formula of the form

$$C_{\gamma\gamma,\ell} \sim C_{\kappa\kappa,\ell} = \frac{9\Omega_M^2 H_0^2}{16} \int_0^\infty dr \frac{g^2(r)}{r^2} P(\ell/r). \tag{2.61}$$

2.3.2.2 Lensing of the Cosmic Microwave Background

Observations of the primary CMB have proven to be one of the most powerful cosmological probes due to the predicted linear evolution of fluctuations from early times till the surface of last scattering. The basic theory of the statistical properties of the CMB at time of emission are fairly well understood and give rise to a characteristic pattern of peaks in the anisotropy power spectra. However, as experiments reach increasing sensitivity nonlinear effects have become increasingly evident. One of the most import of these effects is the gravitational lensing of the cosmic microwave background by structure along the line of sight. This effect is significant both in terms of the cosmological information contains as well as its ability to

affect constraints on fundamental physics (most famously identification of primordial tensor modes). Unlike many other secondary anisotropies, gravitational lensing doesn't have a frequency dependence so it cannot be separated out through multi-frequency observations like the thermal Sunyaev Zeldovich effect or the cosmic infared background.

The unlensed CMB's statistical properties arise from acoustic oscillations in the primordial photon-baryon fluid, where the initial perturbations had a Gaussian and nearly scale invariant spectrum. At $z \sim 10^3$, the universe underwent a phase transition from a ionized plasma to a neutral gas, and the photons are able to free-stream to us. As this combination process is not instantaneous, a photon quadraple moment develops resulting an observed polarization anistropy signal in addition to a temperature anisotropy. As structure forms along the line of sight, this primordial signal gets altered due to gravitational lensing, the integrated Sachs-Wolfe effect, and due to scattering processes.

The magnitude of the lensing effect can be estimated assuming pure Newtonian gravity. Treating the photon like a point mass, one can find the deflection angle as $\beta = 2\phi(R)/c^2$ where $\phi(r)$ is the gravitational potential (relativistic corrections only result in multiplying this by a factor of 2). The depths of these potentials are roughly 10^5 in natural units, resulting in an average deflection of 2×10^{-5} radians. A given photon will likely pass through a number of such potentials between last scattering surface ($\sim$ 14000 Mpc away) and observation. To estimate this number we can look at the peak of matter power spectra, finding that overdensities should be on about $\sim$ 300 Mpc scales. This means the average photon will undergo significant deflections roughly 50 times, resulting in a total angular change of $50 \times 2 \times 10^5 = 10^{-3}$ radians, or $\sim$ 3 arminutes. This scale corresponds to an $\ell \sim 3000$ in terms of CMB multipoles. At these high multipoles Silk damping results in an exponential suppression of the primary CMB, meaning that the resulting structure seen at this high ℓ comes primarily from lensing and other secondary anisotropies.

A more rigorous analysis can be done following the derivation of the expression in Sec. 2.3.2.1 in order to calculate the lensing power spectra caused by the intervening matter. We can use this result to define the expected anistropy in the observed CMB by defining the observed (lensed) power spectra in terms of the primary power spectra as

$$\tilde{C}_\ell^T \approx (1 - \ell^2 R^\phi) C_\ell^T + \int \frac{d^2\boldsymbol{\ell}'}{(2\pi)^2} [\boldsymbol{\ell}' \cdot (\boldsymbol{\ell} - \boldsymbol{\ell}')]^2 C^{\phi\phi}_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|} C_\ell^{T\prime} \tag{2.62}$$

where the deflection angle power is defined as

$$R^\phi = \frac{1}{4} \int \frac{d\ell}{\ell} \ell^4 C_\ell^{\phi\phi}. \tag{2.63}$$

The convolutional term in Eq 2.62 has the effect of smoothing the acoustic peaks; effectively CMB power is both smoothed and moved from large scales to smaller scales.

While here we summarized the effect for temperature anisotropies, a similar argument holds for polarization anisotropies. In that case, one should be careful to do the calculation for both Q and U Stokes parameters (rather than the E field) as the net effect of the lensing will be to move power from E modes to B modes (and also from possible primordial B

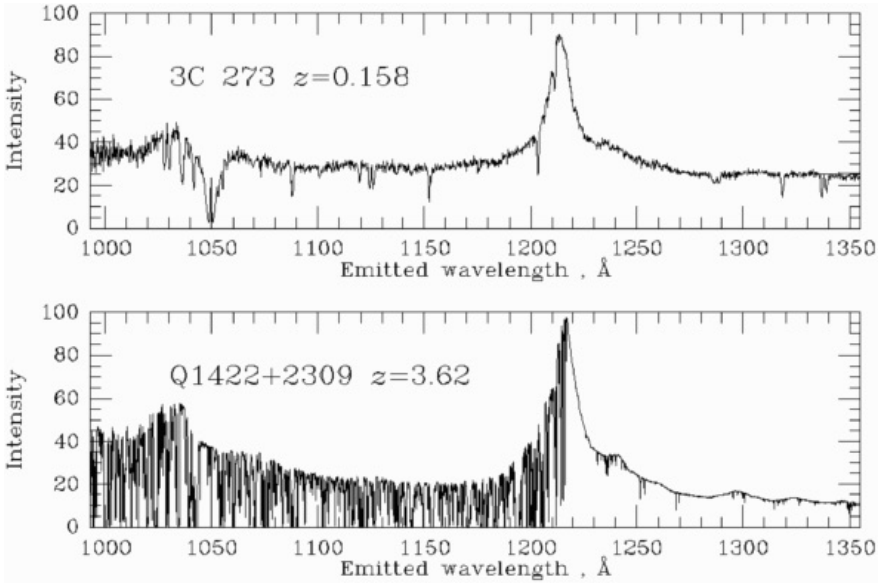


Figure 2.9: *Top:* Rest frame, low redshift quasar spectra without Lyman Alpha Forest. *Bottom:* Rest frame, high redshift quasar spectra with Lyman Alpha Forest present. From `https://pages.astronomy.ua.edu/keel/agn/forest.html`.

modes to E modes). This "leakage" creates a significant systematic for next generation CMB observations trying to detect signatures of inflationary physics. Techniques used to correct this leakage are explored in Chapter 4.

The small scale limit of CMB lensing is particularly interesting as it provides an opportunity to measure the gravitational lensing potential of any astronomical object due to the ubiquity of the CMB on the sky. The physics and detection methods for CMB lensing in this limit are explored in Chapter 3.

2.3.3 Lyman Alpha Forest

The Lyman Alpha forest is a series of absorption features seen in the spectra high redshift quasars and galaxies caused by the presence of neutral hydrogen along the line of sight. As the light travels through a cloud of neutral hydrogen, it rest-frame Lyman alpha frequency (1215.67 Å) gets (partially) absorbed. As there are a number of such features along the line of sight, the result is a series of absorption lines present between the Lyman Alpha and Lyman break feature reminiscent of a forest. This effect was first theorized and observed in 1965 [68], in the spectra of a redshift $z = 2.01$ quasar. This was the first evidence of the presence of a diffuse baryon component to the universe found between galaxies, now commonly known as the Intergalactic Medium (IGM).

The determining quantity in the hydrogen column density N_{HI}. Clouds with extremely high densities ($N_{HI} > 10^{19}$ cm^{-2}) are known as Damped Lyman Alpha Systems (DLAs)

and are characterized by their damping wings on either side of the absorption line. DLAs are usually assumed to be dense, neutral, star forming regions. At lower column densities ($10^{17} < N_{HI} < 10^{19}$) are known as Lyman Limit Systems and exhibit self-shielding; they are optically thick enough that radiation from the surrounding UV background doesn't permeated into their interior. The presence of such systems becomes evident in the far UV (rest frame 987 Å) where they cause a "Lyman break" feature, a discontinuity in the observed spectra. At lower densities ($10^{14} < N_{HI} < 10^{17}$) are the standard diffuse Lyman forest systems who dominate the observed absorption features.

Ignoring presence of metals which could contribute their own signatures, Lyman alpha systems are determined by three parameters; its position in redshift space, its column density, and line width determined by gas temperature and turbulent pressure ($b = \sqrt{2kT/m + b_{turb}^2}$). These parameters determine the entire absorption profile, consisting of a convolution of a Lorentzian resonance curve and a Maxwell-Boltzmann velocity distribution, known jointly as the Voigt profile.

The Lyman Alpha forest is a powerful cosmological probe as, on spatial scales 1 Mpc where the thermal pressure of the gas is not important, the intergalactic medium traces the underlying dark matter mass distribution much more closely than the stellar light of galaxies. This is particularly true at high redshift where observed galaxies would likely be highly biased since only the brightest galaxies can be observed. As long as the gas in the IGM is mostly ionized and in photoionization equilibrium, the optical depth at redshift z can be expressed as;

$$\tau(z) \propto \frac{(\Omega_b h^2)^2}{\Gamma} \frac{H(0)}{hH(Z)} T^{-0.7} \left(\frac{\rho}{\bar{\rho}}\right)^{\alpha} \frac{(1+z)^6}{1 + \frac{dv_{pec}}{H(z)dr}}, \tag{2.64}$$

where Γ is the photoionization rate, $\frac{dv_{pec}}{dr}$ is the gradient of the peculiar velocity along the line of sight, and ρ ($\bar{\rho}$) is the baryon (average) density. This quantity is then further convolved with the Voigt profile and exponentiated ($F = \exp(-\tau)$) to get the observed flux. The value of α depends on the specific gas physics modelled to take into account that denser regions of the universe are typically warmer because it is more effectively heated by photoionization, but will also depend on the reionization history of that particular volume. In the case of an isothermal gas cloud $\alpha = 2$, while for more realistic gas clouds including these complications simulations have found $\alpha \sim 1.5$.

The primary summary statistic used to constrain cosmology using the Lyman Alpha forest is the one dimensional flux power spectra as a function of reshift, $P_F(k, z)$.[39] This statistic is derived from taking the power spectra of each line of sight independently and stacking the derived statistic proportional to the noise properties of each spectra. In order to use this as a cosmological probe there are a number of systematics that need to be corrected for mapping the power spectra to parameter estimates, including hydrodynamical uncertainty in the flux model, redshift space distortions, metal line removal and DLA contamination. While there are some analytical formalism to model some of these effects (for example, [170]), it usually

falls to hydrodynamical simulations to model these effects and compare the estimated power spectrum of those simulations to data.

Another way to make use of Lyman Alpha power spectra is to take advantage of the 3d information; i.e. the spatial correlations between nearby sight-lines. This can be done by computing a three dimension power spectra [61], which should contain strictly more information than its one dimensional counterpart. As the density of lines of sight increases in future Lyman Alpha forest surveys (such as DESI), it becomes more important to measure this statistic due to huge amount of modes available. In practice, however, this statistic is difficult to compute due to explosion of possible pixel pairs in large data-sets. A different way to use this three dimensional structure is to tomographically reconstruct the underlying flux density field, which is discussed in more detail in Chapter 5.

Chapter 3

Reconstructing Small Scale Lenses in the Cosmic Microwave Background

3.1 Introduction

Weak gravitational lensing of the Cosmic Microwave Background (CMB) by the large-scale structure of the Universe produces a signature in the CMB anisotropies in both temperature and polarization and provides a unique probe of the matter distribution out to high redshift ([173]). Lensing induces statistical anisotropies in the observed CMB, through which the mass distribution can be probed. The most common method of lensing reconstruction relies on the quadratic estimator (QE) of Hu and Okamoto ([85]). Previous work ([76, 26]) has shown that, since the QE is only an approximation to the maximum likelihood solution, significant improvements are possible in polarization at low noise levels, usually at the cost of a much greater computational complexity.

In this paper, we revisit lensing reconstruction on very small scales, in a regime where the primary CMB fluctuations have been Silk damped and are negligible (i.e. $\ell 4000$). We will first show that the quadratic estimator can be very suboptimal in this regime as well, since it is limited by the cosmic variance of the long-wavelength background gradient mode, while a true maximum likelihood solution should be able to reconstruct small lenses arbitrarily well, given small enough instrumental noise and residual foreground levels. We then show that the statistical power of the QE is further reduced by the fact that the errors become highly correlated, so that the utility of measuring more modes is reduced when the proper covariance between them is taken into account. Finally, we show that a simple gradient inversion matched filtering approach, as proposed by [172], is close to optimal for reconstruction of very small scale lenses, and that it is not limited by cosmic variance, thus in principle allowing arbitrarily large improvements over the QE (for small enough noise).

One application of this formalism with cosmological significance is cluster mass determination though CMB lensing, or "cluster lensing". The number counts of galaxy clusters are a key prediction of the Λ-CDM model which is sensitive to the overall matter density, Ω_m,

as well as the amplitude of density fluctuations, σ_8, and hence the neutrino mass ΣM_ν. A key limiting factor of using galaxy clusters for cosmological studies is their mass calibration, which is difficult to obtain ([163]). Since it isn't directly observable, cosmologists are left to study different proxies of mass, such as X-ray luminosity or Sunyaev-Zel'dovich signal, and correlate them with gravitational effects of the cluster, such as galaxy-cluster lensing. The uncertainty in the calibration of these relations now limits the constraining power of cluster counts as a cosmological probe ([131]).

While low-redshift massive clusters can be accurately weighed by using lensing of background galaxies along the line of sight ([123, 127]), this approach fails for more distant clusters (at $z1$) because of the lack of enough resolved galaxies on the background. In addition, uncertainties in the photometric redshift distribution of source galaxies, which is challenging to determine precisely at high redshift, can lead to errors in the mass determination.

Meanwhile, the CMB provides a source at known redshift ($z \approx 1100$) and with well studied statistical properties. By looking at the lensing of the CMB by these galaxy clusters, we can avoid some of the shortcomings of galaxy lensing surveys and probe more distant clusters ([78, 212, 213, 136]). While development of the theory behind these measurements dates back almost two decades ([215, 172, 121]), it is only fairly recently that this lensing has been detected at high statistical significance ([136, 128, 15])

These detections have used a modified version of the QE ([84]) which relies on reconstructing the lensing convergence κ, and then applying a matched filter ([136]) on the resulting convergence map. However, as discussed in ([212]), QE obtains unbiased and optimal results only in the limit of no gravitational lensing and becomes progressively more biased and sub-optimal as the gradient in lensing potential increases. Clusters found in SZ samples have particularly strong potential gradients due to their large mass and pose a particularly serious problem for quadratic estimator techniques. The work by ([15]) employs a maximum-likelihood approach and should be immune to the discussion above, but is more computationally expensive.

Revisiting the work of [172] in the context of upcoming CMB experiments such as CMB S4 ([1]), we discuss how an unbiased estimate for the mass of a small spherically symmetric lens (such as an idealized cluster) can be obtained with a simple matched filter of the temperature and polarization maps. In the low-noise regime, we show that these estimates are close to the maximum likelihood solution, and considerably better than the QE. Reconstructing the lensing potential on small scales is particularly important as upcoming CMB experiments can potentially reach sensitivities of ~ 1 μK-arcmin and 1 arcmin beams size. With this increase in instrumental sensitivity, there should be corresponding increases in cluster mass calibration and cosmological constraints ([124, 127]).

In Section 3.2, we briefly review weak lensing of the CMB, explore the small-scale limit of the quadratic estimator and explain the matched filter approach, as well as our assumptions about the cluster mass profile. In Section 3.3, we introduce a matched filter technique we use directly on the temperature map. In Section 3.4, we implement our technique on realistic lensed CMB maps and compare against the quadratic estimator, finding improved performance at low noise levels and high redshifts. In Section 3.5, we compare the results of

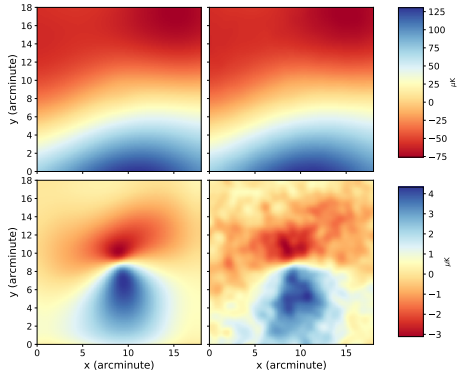


Figure 3.1: Top: Simulated images of CMB lensing caused by a 2×10^{14} $M_\odot$ NFW cluster at $z = 0.7$. Bottom: Unlensed image subtracted out. Left: Beam alone, Right: Beam and 1 μK-arcmin noise.

the matched filter estimate to other lensing reconstruction techniques. Finally, in Section 3.6, we will summarize our results and look at prospects for applying it to future experiments.

For our analysis we will assume the flat ΛCDM Planck 2013 cosmology ([36]), with $H_0 = 67.8$ km (Mpc s)$^{-1}$, and $\Omega_m = 0.307$. We will assume the Born approximation, and ignore field rotation/multiple-lens effects throughout this work.

3.2 Formalism

3.2.1 Effect of Cluster Lensing on the CMB

Here we briefly review the effects of gravitational lensing on the CMB; for a more complete review see ([120]). The observed CMB anisotropies can be described by their temperature fluctuations as a function of direction $\hat{\mathbf{n}}$, $\Theta(\hat{\mathbf{n}})$, as well as two Stokes parameters describing their polarization $Q(\hat{\mathbf{n}})$ and $U(\hat{\mathbf{n}})$. Since the effect of lensing is a simple coordinate remapping, the observed (i.e. lensed) field, $(\tilde{\Theta}, \tilde{Q}, \tilde{U})$, can be related to the primary fields, (Θ, Q, U), through a deflection angle field $\boldsymbol{d}(\hat{\mathbf{n}}) = \nabla\phi$, where ϕ is the lensing potential, by

$$\tilde{\Theta}(\hat{\mathbf{n}}) = \Theta(\hat{\mathbf{n}} + \nabla\phi)) \approx \Theta(\hat{\mathbf{n}}) + \nabla\phi \cdot \nabla\Theta(\hat{\mathbf{n}}) + \dots \tag{3.1a}$$

$$\tilde{Q}(\hat{\mathbf{n}}) = Q(\hat{\mathbf{n}} + \nabla\phi)) \approx Q(\hat{\mathbf{n}}) + \nabla\phi \cdot \nabla Q(\hat{\mathbf{n}}) + \dots \tag{3.1b}$$

$$\tilde{U}(\hat{\mathbf{n}}) = U(\hat{\mathbf{n}} + \nabla\phi)) \approx U(\hat{\mathbf{n}}) + \nabla\phi \cdot \nabla U(\hat{\mathbf{n}}) + \dots \tag{3.1c}$$

Here we have truncated the expansion to first order in $\nabla\phi$, and which is referred to as the "gradient approximation."

Note that while the gradient approximation can be a poor approximation on large scales, it is very good on small enough scales, for the following reason: CMB fluctuations have most of the power on large scales, small scales being suppressed by Silk damping. To quantify which scales contribute to the gradient, we can calculate the variance of the gradient when including multipoles ℓ only up to ℓ_{max}. For temperature, this is

$$G^2_{rms}(<\ell_{\rm max}) = \int_0^{\ell_{\rm max}} \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\, \ell^2 C_\ell^{TT}. \tag{3.2}$$

and a similar definition holds for polarization. Figure 3.2 shows G_{rms} as a function of ℓ_{max}, and we can see that the gradient variance entirely originates at $\ell 2000$, as pointed out in [84]. If we consider small lenses such that $\nabla\phi$ receives most of its contribution from $\ell > 2000$, the gradient approximation should be excellent, and in this paper we will study this regime, which is where analytic progress can be made. For larger lenses, where this assumption fails, more expensive numerical maximum-likelihood methods should be employed to ensure optimality ([76, 26, 137]).

As we have discussed in the introduction, the advantages of CMB lensing over galaxy lensing for measuring cluster masses become large at $z1$, and we will show that most of those clusters are small enough in the sky for the gradient approximation to hold. This makes cluster lensing an ideal application of our formalism.

For concreteness, in the rest of the paper we will work in terms of the temperature fluctuations Θ, but the same formalism also applies to Q and U, since they are deflected by the same vector. Sometimes it will also be useful to work in terms of the lensing convergence, κ, defined as $\kappa(\hat{\mathbf{n}}) = -\frac{1}{2}\nabla\cdot\boldsymbol{d}(\hat{\mathbf{n}}) = -\frac{1}{2}\nabla^2\phi(\hat{\mathbf{n}})$. We also note, that since the rms gradient of temperature fluctuations is almost an order of magnitude larger than that for polarization, temperature reconstruction is expected to be the dominant source of information on about small scales. Polarization, while statistically less powerful, is less affected by some of the foreground contamination and can provide useful consistency checks.

3.2.2 Quadratic Estimator in the small-scale limit and Optimal Estimators

CMB lensing introduces coupling between long and short wavelength modes and it is possible to construct a minimum-variance quadratic estimator for the lensing potential ϕ. Note that the "minimum-variance" qualification here only applies to the class of estimators that are quadratic in the observed fluctuations, and is not a general statement of optimality. It has been previously shown that maximum likelihood or iterative methods can outperform the QE. In this section we show that the QE is very suboptimal in the small-scale and low noise regime since it is subject to the cosmic variance on the gradient mode, while (in this regime) the particular realization of the gradient on the background of the lens can be measured

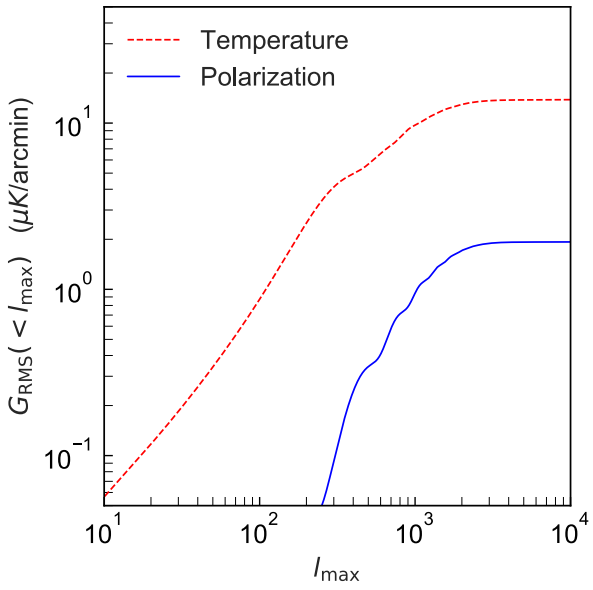


Figure 3.2: Gradient of the lensed temperature T and polarization E fields as a function of the filtering scale, ℓ_{max}, as defined in Eq 3.2. We note that the RMS gradient is almost a factor of 10 larger in temperature than in polarization, and this will make temperature lensing more sensitive to a fixed-mass lens than polarization on small scales (in absence of foregrounds).

without cosmic variance. In this section we will take the limit in which the reconstructed mode $L \gg \ell_{Silk}$ (in practice $L4000$), low noise and no-foreground limit.

The standard QE of Hu and Okamoto ([85]) for temperature can be written as[1]:

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) = N(\boldsymbol{L}) \int_{\boldsymbol{\ell}} \tilde{\Theta}(\boldsymbol{\ell})\tilde{\Theta}(\boldsymbol{L}-\boldsymbol{\ell}) f(\boldsymbol{\ell},\boldsymbol{L}) \tag{3.3}$$

where the mode-coupling kernel is

$$f(\boldsymbol{\ell},\boldsymbol{L}) = \frac{(\boldsymbol{L}-\boldsymbol{\ell})\cdot\boldsymbol{L} C^{TT}_{|\boldsymbol{L}-\boldsymbol{\ell}|} + \boldsymbol{\ell}\cdot\boldsymbol{L} C^{TT}_{\ell}}{2C^{\rm tot}_{\ell} C^{\rm tot}_{|\boldsymbol{L}-\boldsymbol{\ell}|}} \quad . \tag{3.4}$$

Here, C^{tot}_{ℓ} includes contributions from instrumental noise, foregrounds, and the (lensed) primary CMB. The reconstruction noise serves as the normalization in the estimator and represents the uncertainty in the reconstruction of $\phi(\boldsymbol{L})$ due to chance correlations between

[1]For compactness, we use the notation $\int_{\boldsymbol{\ell}} \equiv \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}$. Upper-case $\mathbf{L}$ denotes lensing multipole, while lower-case $\boldsymbol{\ell}$ denotes temperature map multipole.

different modes in an unlensed, Gaussian realization,

$$N(\boldsymbol{L})^{-1} = \int_{\boldsymbol{\ell}} \frac{\left[(\boldsymbol{L}-\boldsymbol{\ell})\cdot\boldsymbol{L}C^{TT}_{|\boldsymbol{L}-\boldsymbol{\ell}|} + \boldsymbol{\ell}\cdot\boldsymbol{L}C^{TT}_{\ell}\right]^2}{2C^{tot}_{\ell}C^{tot}_{|\boldsymbol{L}-\boldsymbol{\ell}|}} \quad . \tag{3.5}$$

Taking the expectation value of this estimator over many CMB realizations for fixed ϕ, this estimator recovers an unbiased mean deflection field (at first order in the lensing expansion). However, this reconstruction has significant noise for any given CMB realization. This can be seen by taking the limit of reconstruction of small scale modes, with L higher than where the primary CMB C^{TT}_{ℓ} has support, i.e. on scales much smaller than the Silk damping scale, $L \gg \ell_{Silk}$. Since we will use L for the reconstructed multipole and ℓ for the multipoles used in the reconstruction, we'll denote this limit by $L \gg \ell$. We then have:

$$f(\boldsymbol{\ell},\boldsymbol{L}) \xrightarrow{L\gg\ell} 2\,\frac{\boldsymbol{\ell}\cdot\boldsymbol{L}C^{TT}_{\ell}}{2C^{tot}_{\ell}C^{tot}_{L}}, \tag{3.6}$$

and

$$N(\boldsymbol{L})^{-1} \xrightarrow{L\gg\ell} 2\int_{\boldsymbol{\ell}} \frac{\left[\boldsymbol{\ell}\cdot\boldsymbol{L}C^{TT}_{\ell}\right]^2}{2C^{tot}_{\ell}C^{tot}_{L}}. \tag{3.7}$$

Therefore the quadratic estimator simplifies to[2]

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) \longrightarrow \frac{\displaystyle\int_{\boldsymbol{\ell}} \boldsymbol{\ell}\cdot\boldsymbol{L}\; W_F(\ell)\tilde{\Theta}(\boldsymbol{\ell})\tilde{\Theta}(\boldsymbol{L})}{\displaystyle\int_{\boldsymbol{\ell}} (\boldsymbol{\ell}\cdot\boldsymbol{L})^2\; W_F(\ell)C^{TT}_{\ell}} \;, \qquad with\; W_F(\ell) = \frac{C^{TT}_{\ell}}{C^{tot}_{\ell}} \;. \tag{3.8}$$

In the low-noise, no foregrounds limit, we can take the Wiener filter $W_F(\ell) \to 1$ and note that

$$\int_{\boldsymbol{\ell}} (\boldsymbol{\ell}\cdot\boldsymbol{L})^2\; W_F(\ell)C^{TT}_{\ell} \approx \int_{\boldsymbol{\ell}} (\boldsymbol{\ell}\cdot\boldsymbol{L})^2\; C^{TT}_{\ell} = L^2\;\frac{1}{2}(\nabla\Theta)^2_{rms} \;. \tag{3.9}$$

The quadratic estimator takes a simpler form of

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) \longrightarrow \frac{\tilde{\Theta}(\boldsymbol{L})}{L}\;\frac{\displaystyle\int_{\boldsymbol{\ell}} \boldsymbol{\ell}\cdot\hat{\mathbf{n}}_{\boldsymbol{L}}\;\tilde{\Theta}(\boldsymbol{\ell})}{\frac{1}{2}(\nabla\Theta)^2_{rms}} \;, \tag{3.10}$$

[2]Technically, the convergence in the following equation would only be "in probability" and not at the field level, if $\Theta(\boldsymbol{\ell})$ and $\phi(\boldsymbol{\ell})$ were random variables with values that are uncorrelated for different $\boldsymbol{\ell}$. In practice, the finite (and small) window function used to define the local gradient and nonlinear evolution of the potential make both continuous functions and the limit well defined. This subtlety doesn't affect any of our results.

where we have written $\boldsymbol{L} = L\ \hat{\mathbf{n}}_{\boldsymbol{L}}$. Here $\int_{\boldsymbol{\ell}} \boldsymbol{\ell}\cdot\hat{\mathbf{n}}_{\boldsymbol{L}}\ \tilde{\Theta}(\boldsymbol{\ell}) \approx (\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}$ is the (realization dependent) measured background gradient projected in the direction parallel to $\boldsymbol{L}$. We can then rewrite

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) \longrightarrow \frac{\tilde{\Theta}(\boldsymbol{L})}{L}\ \frac{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}}\ . \tag{3.11}$$

This is the small-scale limit of the QE, which we will interpret soon.

On the other hand, at $L > 4000$, the primary CMB is highly suppressed by diffusion damping and in our limit, all small scale fluctuation are created by lensing. In this regime, we can write the small-scale fluctuations as

$$\tilde{\Theta}(\hat{\mathbf{n}}) \approx \nabla\phi\cdot\nabla\Theta(\hat{\mathbf{n}})\ . \tag{3.12}$$

Note that we can treat the locally measured gradient as constant on scales smaller than $1/\ell_G$, where $\ell_G \approx 2000$ is the multipole where the gradient becomes saturated as shown in figure 3.2. Then, there is a one-to-one correspondence between small scale lenses $\nabla\phi$ and measured fluctuations $\tilde{\Theta}$. In Fourier space,

$$\begin{aligned}\tilde{\Theta}(\boldsymbol{L}) &= \int_{\boldsymbol{\ell}} \boldsymbol{\ell}\cdot(\boldsymbol{L}-\boldsymbol{\ell})\Theta(\boldsymbol{\ell})\phi(\boldsymbol{L}-\boldsymbol{\ell}) \\ &\xrightarrow{L\gg\ell}\ L\phi(\boldsymbol{L})\int_{\boldsymbol{\ell}} \boldsymbol{\ell}\cdot\hat{\mathbf{n}}_{\boldsymbol{L}}\ \tilde{\Theta}(\boldsymbol{\ell}) = L\phi(\boldsymbol{L})\ (\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}\ .\end{aligned} \tag{3.13}$$

We can call the solution of the previous equation for $\phi(\boldsymbol{L})$ the "Gradient Inversion" (GI) solution:[3]

$$\hat{\phi}_{\rm GI}(\boldsymbol{L}) \longrightarrow \frac{\tilde{\Theta}(\boldsymbol{L})}{L}\ \frac{1}{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}\ . \tag{3.14}$$

Note that modes with $\boldsymbol{L}$ perpendicular to the gradient direction have $(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}} = 0$ and cannot be reconstructed. This is because in this limit, lensing doesn't produce any effect perpendicular to the gradient direction and therefore no estimator can reconstruct such modes.

We can now compare the two estimators:

$$\begin{aligned}L\ \hat{\phi}_{\rm QE}(\boldsymbol{L}) &\longrightarrow \tilde{\Theta}(\boldsymbol{L})\ \frac{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}}, \\ L\ \hat{\phi}_{\rm GI}(\boldsymbol{L}) &\longrightarrow \tilde{\Theta}(\boldsymbol{L})\ \frac{1}{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}\ .\end{aligned} \tag{3.15}$$

Given a good enough experiment, $\tilde{\Theta}(\boldsymbol{L})$ and $(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}$ can be made measured to arbitrary accuracy, and therefore the error on $\hat{\phi}_{\rm GI}(\boldsymbol{L})$ can be made arbitrarily small. Note that in

[3]The GI solution corresponds to the maximum likelihood solution in the limit considered here, where both the gradient and the short scale modes can be measured on this particular realization with $S/N \gg 1$. This can be seen by writing a likelihood based on Equation 3.12 and taking the noise to zero.

this limit neither $\tilde{\Theta}(\boldsymbol{L})$ nor $(\nabla\Theta)_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}$ are random variables, and should therefore not be marginalized over.

This is not the case for the quadratic estimator, which can be rewritten as (in the limit of small noise), and identifying $\hat{\phi}_{\rm GI}$ with the "true" lensing potential ϕ_{true},

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) = \hat{\phi}_{\rm GI}(\boldsymbol{L})\ \frac{(\nabla\Theta)^2_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}} \longrightarrow \phi_{true}(\boldsymbol{L})\ \frac{(\nabla\Theta)^2_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}} \tag{3.16}$$

Firstly, we notice that $\langle(\nabla\Theta)^2_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}\rangle = \frac{1}{2}(\nabla\Theta)^2_{rms}$, so that $\langle\hat{\phi}_{\rm QE}\rangle = \phi_{true}$, where the expectation value is taken over realizations of the background gradient, for a fixed ϕ. We see that while unbiased, the "error" that the QE makes is proportional to the difference between the (square) true gradient and the rms gradient. Since this quantity is on average of order the rms gradient itself, the fractional error of the QE is always of order unity per mode.

We can formalize the above intuition, showing that even with arbitrarily small experimental noise, there is a lower limit to the statistical error on the quadratic estimator $\sigma(\hat{\phi}_{\rm QE})$. Defining

$$R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}}) = \frac{(\nabla\Theta)^2_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}} \quad , \tag{3.17}$$

we find that

$$\left(\frac{\sigma^2(\hat{\phi}_{\rm QE})}{\phi^2_{true}}\right)_{min}(\boldsymbol{L}) = \left\langle(R^2_\nabla)^2\right\rangle - \left\langle R^2_\nabla\right\rangle^2 = 3 - 1 = 2 \tag{3.18}$$

In summary, the GI estimator can have arbitrarily low noise per mode for a good enough experiment, while the QE is limited to $S/N = 1/\sqrt{2}$ per mode (which is the same has having cosmic variance on the gradient mode).

3.2.3 One gradient, many independent modes?

In this section we explore the covariance of the small scale lensing modes estimated using the quadratic estimator, in the very high-L, very low noise limit. In this regime, we have shown that

$$\hat{\phi}_{\rm QE}(\boldsymbol{L}) \longrightarrow \phi_{true}(\boldsymbol{L})\ \frac{(\nabla\Theta)^2_{true,\hat{\mathbf{n}}_{\boldsymbol{L}}}}{\frac{1}{2}(\nabla\Theta)^2_{rms}} = \phi_{true}(\boldsymbol{L})\ R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}}) \tag{3.19}$$

Then the noise covariance between different small scale modes sharing the same background gradient becomes:

$$N^{\phi\phi}_{\rm QE}(\boldsymbol{L},\boldsymbol{L}') \equiv \langle(\hat{\phi}_{\rm QE}(\boldsymbol{L}) - \phi_{true}(\boldsymbol{L}))\ (\hat{\phi}_{\rm QE}(\boldsymbol{L}') - \phi_{true}(\boldsymbol{L}'))\rangle_\Theta \tag{3.20}$$

where crucially the average is taken over realization of the long-wavelength CMB fluctuations, for a fixed large scale structure ϕ_{true}. we have

$$\begin{aligned} N^{\phi\phi}_{\rm QE}(\boldsymbol{L},\boldsymbol{L}') &\longrightarrow \phi_{true}(\boldsymbol{L})\phi_{true}(\boldsymbol{L}')\langle(R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}}) - 1)(R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}'}) - 1)\rangle \\ &= 2\cos^2(\zeta)\ \phi_{true}(\boldsymbol{L})\phi_{true}(\boldsymbol{L}') \end{aligned} \tag{3.21}$$

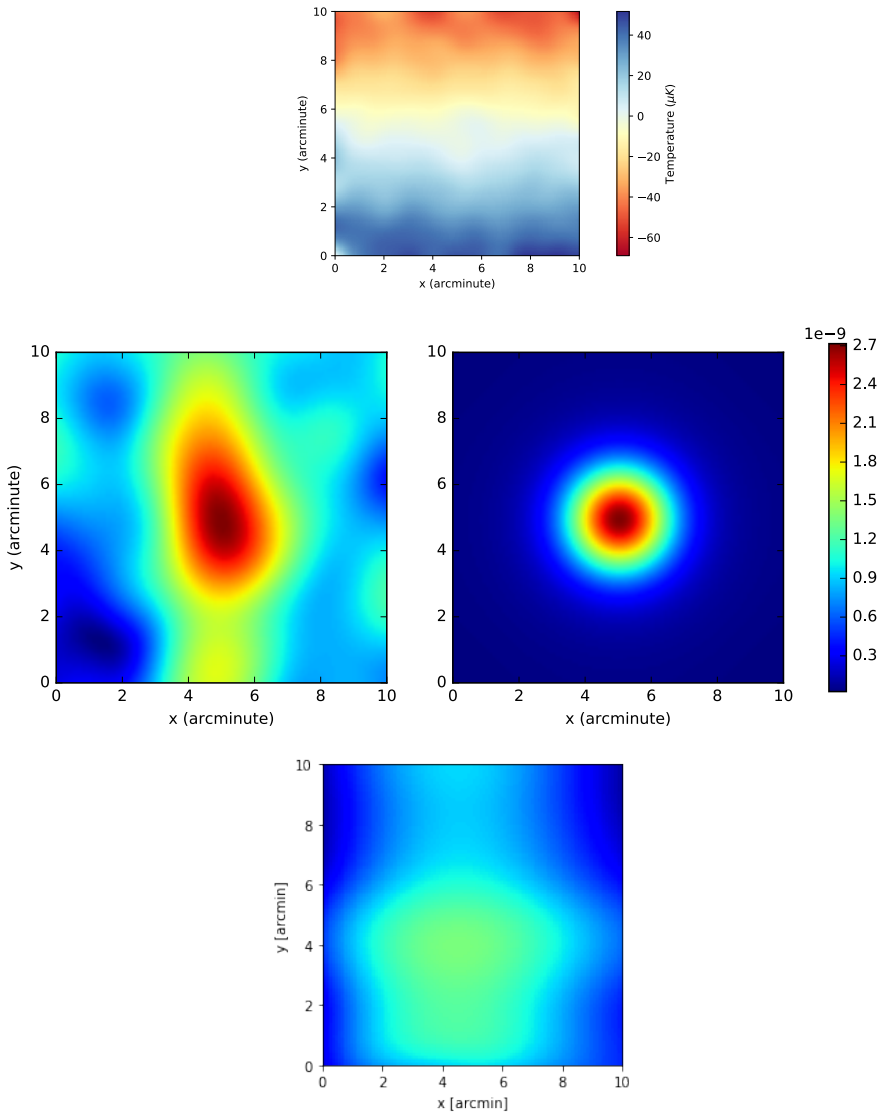


Figure 3.3: Top: Simulated lensed CMB Map of a constant gradient in the $\hat{y}$ direction, assuming 1 arcmin beam and noise level of 1 μK-arcmin. A cluster mass of $2 \times 10^{14} M_{\odot}$ at $z = 2.0$ with an NFW profile (see Section 3.3.1) is present at the center. Middle Left: Reconstruction of the convergence map given the observed CMB map using the GI method. Note that pure vertical modes (i.e. mass density with only $\hat{y}$ component) have been explicitly set to zero, while modes close to pure $\hat{y}$ are poorly reconstructed resulting in a residual vertical band. Middle Right: True convergence of cluster. Bottom: Quadratic estimator convergence of cluster at same color-scale.

where ζ is the angle between $\boldsymbol{L}$ and $\boldsymbol{L}'$ and we have used $\left\langle R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}}) R^2_\nabla(\hat{\mathbf{n}}_{\boldsymbol{L}'})\right\rangle = 1 + 2\cos^2(\zeta)$ and $\langle R^2_\nabla\rangle = 1$.

We conclude that for small scale modes not only the noise covariance matrix $N^{\phi\phi}_{QE}(\boldsymbol{L}, \boldsymbol{L}')$ is not diagonal, but the off-diagonal terms become as large as the on-diagonal ones. This means that when detector noise is negligible, modes sharing the same gradient become highly correlated, and treating them as independent (for example when forecasting cluster lensing constraints for future surveys), may lead to underestimation of the overall statistical noise.

We also note that this is a feature of the quadratic estimator, and won't be a limiting factor in a maximum-likelihood approach that includes information about the large-scale gradient. In principle, a maximum likelihood approach has no limit on the statistical significance that it can achieve[4].

3.2.4 Example: Lensing Map Reconstruction

In the previous section, we have derived a ϕ estimator for the pure gradient, low noise limit, which is found in Eq 3.14. Equivalently, we can rephrase that in terms of the lensing convergence $\kappa(\boldsymbol{L}) = -\frac{L^2}{2}\phi(\boldsymbol{L})$,

$$\hat{\kappa}_{\mathrm{GI}}(\boldsymbol{L}) \longrightarrow -\frac{L\tilde{\Theta}(\boldsymbol{L})}{2}\ \frac{1}{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}\ . \tag{3.22}$$

In Figure 3.3, we show a comparison between the true and reconstructed convergence field, in the case of a massive lens with mass $2\times10^{14}M_\odot$ and following an NFW profile, as described in the next section. We note that as expected, the modes with variation perpendicular to the gradient direction are not correctly reconstructed, explaining the vertical band in the reconstructed map. Those modes can be filtered out or downweighted in a real analysis since they have infinite variance.

In presence of finite noise, the error on the estimate of a given $\kappa(\boldsymbol{L})$ noise, depends on the errors on both the large-scale gradient $(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}$ and the small scale temperature fluctuation $\tilde{\Theta}(\boldsymbol{L})$.

We expect the error on the gradient to be almost negligible in most cases, unless the gradient on the patch of interest happens to be much smaller than the rms, since most current or future CMB experiments should be close to cosmic variance limited at $\ell 2000$,

$$\left(\frac{\sigma(\hat{\kappa}_{\mathrm{GI}})}{\kappa}\right)^2 = \left(\frac{\sigma(\tilde{\Theta})}{\tilde{\Theta}}\right)^2 + \left(\frac{\sigma((\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}})}{(\nabla\Theta)_{true,\hat{\mathbf{n}}\boldsymbol{L}}}\right)^2 . \tag{3.23}$$

Assuming that the errors are on the gradient and the small scales are independent. Note that the noise properties of $\hat{\kappa}_{GI}$ are in general not trivial, since the noise is anisotropic (it depends

[4]Further neglecting post-Born corrections and field rotation, as well as foregrounds and residual primary CMB.

on the angle between the gradient and $\boldsymbol{L}$), and the direction of the anisotropy depends on position on the sky. We defer a full treatment to future work, but we will discuss noise estimates in the specific case of cluster lensing in what follows.

3.3 Measuring Cluster Masses with matched filtering

A straightforward application of the formalism outlined above arises when the lens profile is known up to a normalization factor, such as when determining the mass M of a cluster, lying on a constant gradient $\boldsymbol{G} = (\nabla\Theta)_{true}$. We have previously shown that the on small scales, the GI solution is linear in the measured temperature fluctuations, and therefore the minimum variance estimator for the overall amplitude of the lensing $A = M|G|$, can be obtained by a matched filter of the CMB map ([71]). Once both the amplitude A and the background gradient $\boldsymbol{G}$ have been measured, the mass can be obtained by direct inversion $M \approx A/|G|$. In the context of cluster lensing, this was first investigated by ([172]), and here we revisit this point.

A mass of a given profile on a pure gradient background CMB will create a dipole-like structure $g(\boldsymbol{\theta})$, aligned with the gradient direction, and with a free amplitude A that we wish to measure. We then model the total small-scale CMB anisotropy as

$$\tilde{\Theta}(\boldsymbol{\theta}) = Ag(\boldsymbol{\theta}) + n(\boldsymbol{\theta}) \tag{3.24}$$

where $g(\boldsymbol{\theta})$ is the angular profile of the deflection angle of a known profile (i.e. NFW, Sersic, etc.) caused by the cluster, A is an amplitude depending on the mass of the cluster and the CMB gradient, and $n(\boldsymbol{\theta})$ is noise, either from the instrument, foregrounds, or residual primary CMB. It is useful to work with the noise in Fourier space and its power spectrum,

$$\langle n(\boldsymbol{\ell})\ n(\boldsymbol{\ell}')\rangle = (2\pi)^2 C_\ell^{tot}\delta(\boldsymbol{\ell}+\boldsymbol{\ell}'). \tag{3.25}$$

The value of A can be found by applying a linear filter Ψ on the temperature map at the cluster position

$$\hat{A} = \int d^2\boldsymbol{\theta}\ \Psi(\boldsymbol{\theta})\tilde{\Theta}(\boldsymbol{\theta}), \tag{3.26}$$

where the optimal filtering function is given by ([71]),

$$\Psi(\boldsymbol{\ell}) = \left[\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\frac{|g(\boldsymbol{\ell})|^2}{C_{\boldsymbol{\ell}}^{tot}}\right]^{-1}\frac{g(\boldsymbol{\ell})}{C_\ell^{tot}}, \tag{3.27}$$

can be found by minimizing the error on $\hat{A}$. We show components of our matched filter for a NFW profile in Figure 3.4.

The variance on A is given by

$$\sigma_A^2 = \left[\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\frac{|g(\boldsymbol{\ell})|^2}{C_{\boldsymbol{\ell}}^{tot}}\right]^{-1} \tag{3.28}$$

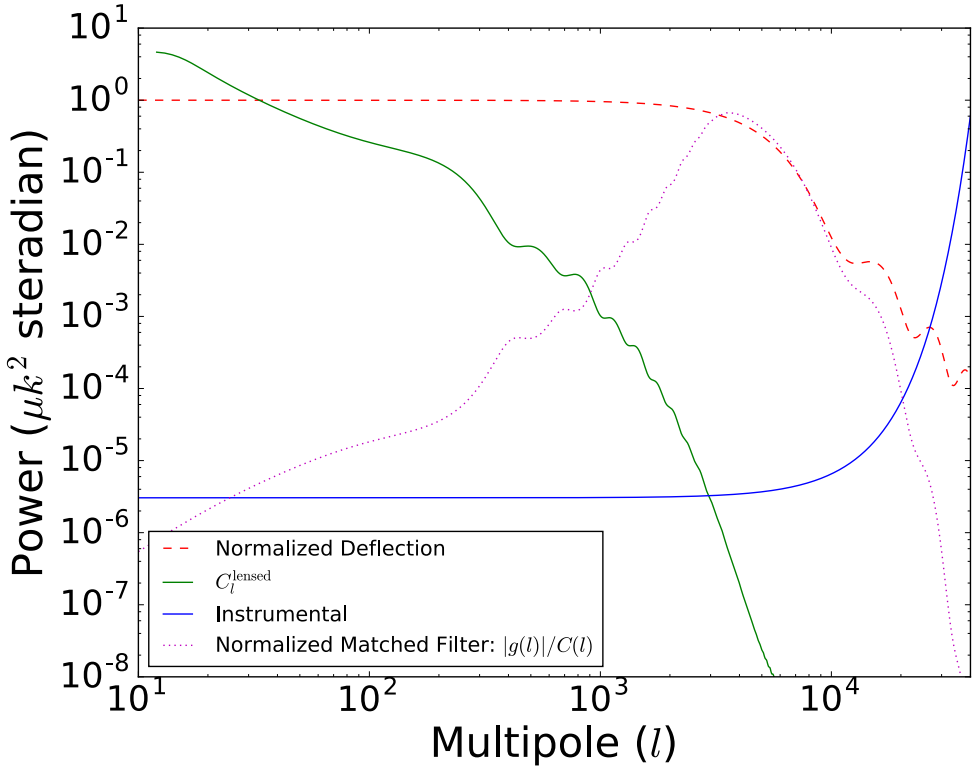


Figure 3.4: Components of matched filter in $\boldsymbol{\ell}$-space. At low $\boldsymbol{\ell}$ the cluster power will be suppressed by the primary CMB fluctuations and at high $\boldsymbol{\ell}$ it will be suppressed by the instrumental beam. Instrumental noise is for a 6 μK-arcmin experiment with a 1 arcmin beam.

Assuming an axi-symmetric cluster profile we can write

$$g(\boldsymbol{\ell}) = g(\ell)\cos\phi_\ell \tag{3.29}$$

where ϕ_ℓ is the angle between the vector $\boldsymbol{\ell}$ and the gradient direction.

Note that $\sigma_A \to 0$ as the noise power spectrum tends to zero, which is what we expect.

3.3.1 NFW Profile

For our work, we will assume that halos follow a NFW profile [141], with density given by

$$\rho(r) \propto \frac{1}{r/r_s(1+r/r_s)^2}, \tag{3.30}$$

with scale radius r_s and normalization constant dependent on halo mass. The halo mass is related to the scale radius via the concentration parameter, c, which we will fix to 3.2 to

allow direct comparison against [84]. In practice, variation of the concentration and scale radius can be corrected for by combining CMB lensing information with other proxies for density, such as tSZ and X-ray observations. The lensing deflection profile (with the speed of light equal to one) for an NFW halo is

$$|\nabla\phi(\theta)| \propto g(\theta) = 16\pi GM\frac{D_{LS}D_L}{D_S}\frac{\rho_s r_s\theta_s}{\theta}h\left(\frac{\theta}{\theta_s}\right) \tag{3.31}$$

where,

$$\rho_s = \frac{200}{3}\rho_{crit}\frac{c^3}{\ln c(1+c) - c/(1+c)} \tag{3.32}$$

and D_L is the comoving distance to the lensing, D_S is the comoving distance to the source (i.e. the CMB), D_{LS} is the comoving distance between the lens and the source, θ_s is the angle subtended by the scale radius and M is the cluster mass. The functional dependence of the profile is

$$h(x) = \begin{cases} ln(x/2) + \frac{2}{\sqrt{x^2-1}}arctan\sqrt{\frac{x-1}{x+1}}, & (x > 1) \\ ln(x/2) + \frac{2}{\sqrt{1-x^2}}arctan\sqrt{\frac{1-x}{1+x}}, & (x < 1) \\ ln(x/2) + 1\,, & (x = 1)\,. \end{cases} \tag{3.33}$$

3.4 Idealized Example on CMB Map

To demonstrate this technique, we generated CMB realizations using HEALPix ([65][5]), with $N_{side} = 2048$, from which we extracted $20'$ square cutouts. For these images we use the lensed CMB power spectra to include the effects of lensing by large scale structure. The effect of lensing from an NFW profile with a given concentration and mass is then added, together with detector noise, as shown in Figure 3.1. We then reconstruct the mass using the matched filter technique described above, the first step being a measurement of the average the background gradient, as discussed below.

3.4.1 Reconstruction of the background gradient

Correct determination of the cluster mass depends on being able to accurately extract the mean background gradient [78], without bias from the cluster lensing signal or deviation from the pure gradient approximation. Fortunately, as discussed in section 3.2.1, this gradient has little variation on cluster scales. We can define the average gradient on a small patch centered on the cluster i as $\boldsymbol{G}_i = [(\nabla\Theta)_{true}]_{patch\ i}$, so that

$$\boldsymbol{G}_i = \frac{1}{C}\int_{\ell<2000}\frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\,[\boldsymbol{\ell}\;\tilde{\Theta}(\boldsymbol{\ell})]_{patch\ i} \tag{3.34}$$

[5]http://healpix.sourceforge.net

where C is proportional to the area being integrated over. One should be careful to not average over too large of a patch such that variations in the primary CMB would become a concern. For our analysis we find that a $8'$ box around the center of the cluster works adequately.

This measure is robust even in the presence of noise, resulting in less than a 0.1% error in gradient extraction for an experiment with a 1 arcmin beam and 1 μK-arcmin sensitivity, compared to the ideal case. The error scales roughly linearly with instrumental sensitivity and has very small dependence on lensing halo mass and redshift.

In addition, the presence of clusters does not strongly bias the measurement of the gradient itself. This shown explicitly in Figure 3.5, where we simulated the lensing of the CMB by massive clusters, extracted the gradient from those lensed images, and applied the matched filter prescription to measure the mass. As shown, there is no significant bias in the measured mass as it relates to the real mass.

3.4.2 Measuring the lens mass

Here we briefly summarize the procedure used for estimating the mass of simulated clusters. For a given simulated lensed image, we perform the following procedure:

1. Find the direction and magnitude of the gradient $\boldsymbol{G}_i$ at the center of the cluster, as in Equation 3.34.

2. Choose the axis of the matched filter (i.e. the $\phi_\ell = 0$ direction), aligned with the direction of the gradient. This is important due to the $\cos\phi_\ell$ term in Equation 3.29 establishing the antisymmetry of the filter in real space.

3. Apply the matched filter defined in Equation 3.27 and perform the integral in Equation 3.26 to obtain an amplitude $\hat{A}_i$.

4. Since $A_i = |G_i|M_i$ when correctly normalized, the mass M_i can be estimated as $\hat{M}_i = \hat{A}_i/|G_i|$, with error dependent on the local size of the gradient as expected: $\sigma(M_i) = \sigma_A/|G_i|$. For a sample of many different clusters, the mean mass can be obtained by inverse noise weighting each measurement as explained in the next section.

3.4.3 Gradient-based weighting of samples

Assume that we have a collection of N clusters all of the same mass, M, but each in their own background gradient, G_i, which we assume is extracted with no noise from observations. Then the matched filter output for each cluster i, $\hat{A}_i$ (from Equation 3.26), and it's uncertainty, σ_A (which is approximately independent of the cluster mass in a uniform sample, and

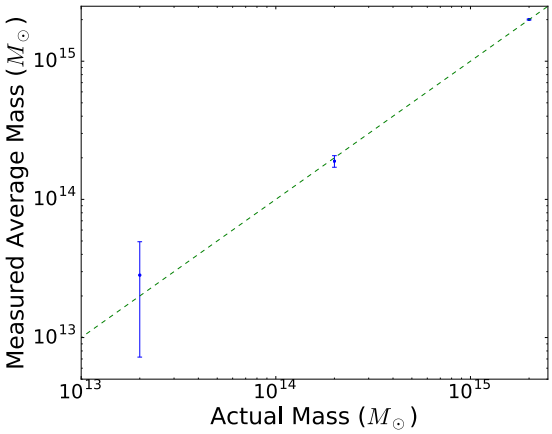


Figure 3.5: Cluster mass reconstruction as a function of mass for our 1000 realistic lensed CMB realizations. We have fixed $z = 0.7$, $c = 3.2$, $\Delta_T = 6\mu$K-arcmin, and beam FWHM = 1 arcmin. Our reconstructed mass has no measurable bias.

only determined by detector noise and foregrounds), are given by

$$\hat{A}_i = |G_i| M_i, \tag{3.35}$$

$$\sigma_A = |G_i| \sigma(M_i). \tag{3.36}$$

The minimum variance estimate of the mean mass is obtained by inverse noise weighing individual measurements. Noting that from Equation 3.36, $\sigma(M_i)$ is inversely proportional to $|G_i|$, we have

$$\hat{M} = \frac{\sum_i \frac{|G_i|\hat{A}_i}{\sigma_A^2}}{\sum_i \frac{|G_i|^2}{\sigma_A^2}}. \tag{3.37}$$

Where the uncertainty on the mean mass is

$$\sigma^2(\hat{M}) = \left[\sum_i \frac{|G_i|^2}{\sigma_A^2} \right]^{-1} = \frac{1}{N} \frac{\sigma_A^2}{G_{rms}^2} \quad , \tag{3.38}$$

The latter equality being true in the limit of a large sample size. Equation 3.38, together with the matched filter error σ_A given by Equation 3.28, provides the basis for our results in Figure 3.6. Note that uniform weighting would lead to an increase of the mass uncertainty.

3.4.4 Results

In Figure 3.6 we show the analytical calculation of the mass uncertainty using the formalism of Equation 3.28, allowing the integral to be cut off by the beam (see Figure 3.4). Our

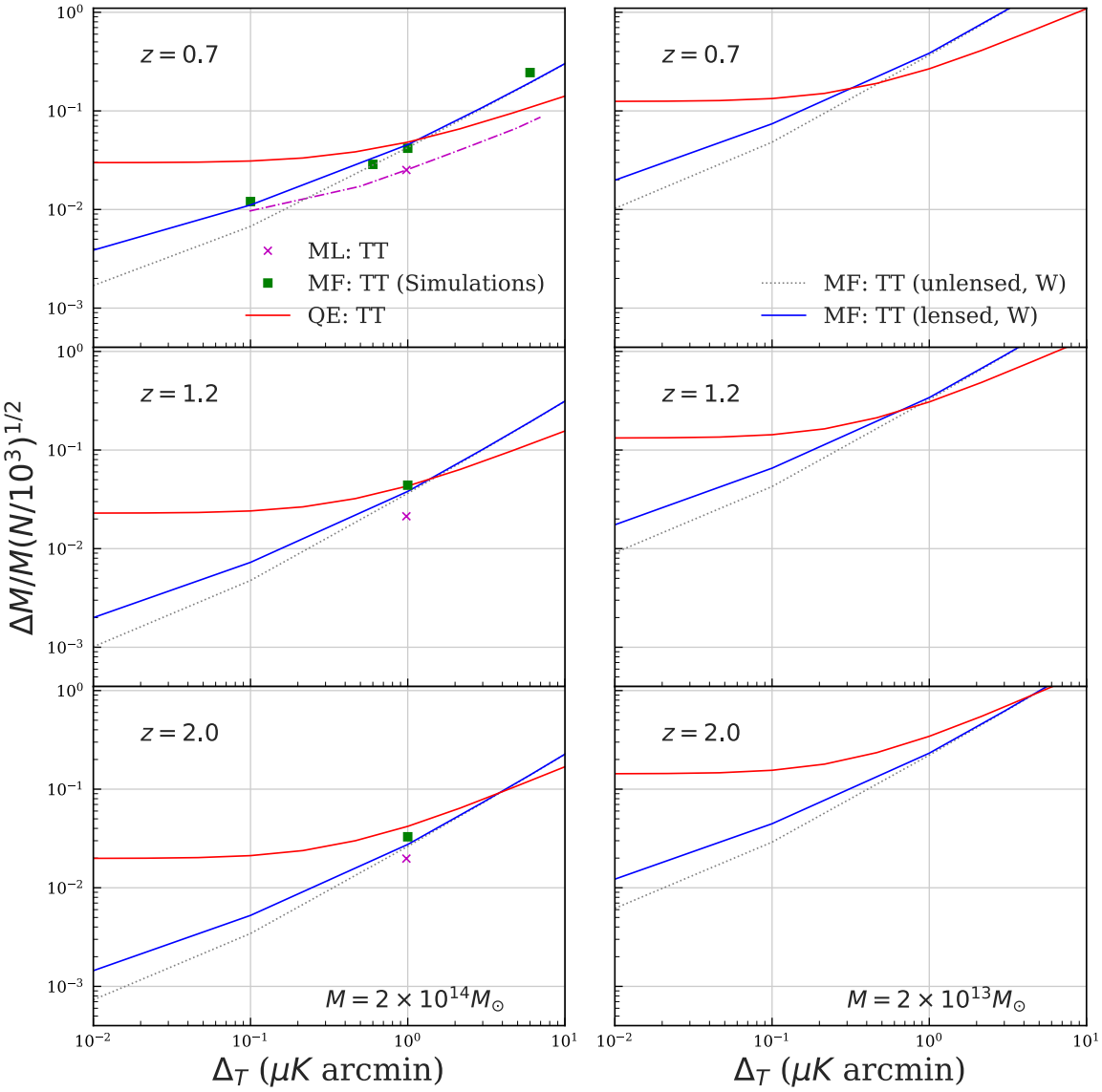


Figure 3.6: Mass sensitivity assuming 1 arcmin beam and no foregrounds using different techniques for redshifts $z = 0.3, 0.7$, and 1.2, as well as for two different masses, $2 \times 10^{14} M_\odot$ (left) and $2 \times 10^{13} M_\odot$ (right). In addition to the matched filter result from this work (MF) we show the quadratic estimator (QE) from [84], which has an essentially flat redshift dependence at $z > 0.5$ for a 1 arcmin beam (see also Figure 3 in [136]). We have also shown the maximum likelihood (ML) TT result from [158] (see Figure 2). In the middle panel, we show the effect of using the lensed vs unlensed CMB as a source of noise. For all $\boldsymbol{\ell}$ integrals we use an $\boldsymbol{\ell}_{min} = 10$ and $\boldsymbol{\ell}_{max} = 40,000$ (beyond the beam cutoff scale).

fiducial case include the residual *lensed* CMB as a source of noise in C_ℓ^{tot}, while we also show the case in which the unlensed CMB is used instead. In practice, the fiducial case is conservative in the sense that no delensing is assumed, while any amount of delensing (either internal ([27]) or with the use of other tracers ([184])) will place the noise between our two extreme cases.

For comparison, we show the performance of the quadratic estimator (QE) of [84], as well as the maximum likelihood (ML) result of [158]. For the QE, we use the formalism of [84] (see Equation 30 there) and we integrate over ℓ well past the beam cutoff scale ($\ell \approx 10,000$) for a fair comparison between the methods. The matched filter outperforms the quadratic estimator in the low noise limit as well as the high redshift limit. For a cluster of a constant mass, its apparent size decreases with increasing redshift and the background gradient is closer to a pure gradient where the match filter technique is optimal.

As expected by the analytical calculations in Section 3.2.2, the quadratic estimator performance saturates at signal to noise ratio $S/N \sim 1$ per cluster. Meanwhile, the matched filter keeps improving for small values of Δ_T.

As a check, we validate our error estimate on simulations. To do this, we have taken 1,000 cutouts from full-sky realizations of the lensed primary CMB, as described in Section 3.4 and placed an NFW cluster with constant mass and $c = 3.2$ at the center of the image and lensed the background primary CMB. We then measured the gradient as described in Section 3.4.1, and found their mass using the procedure outlined in Section 3.4.2. We then weight the clusters according to their background gradient, as described in 3.4.3, to find the average measured mass; the results of which are shown in Figure 3.6.

3.4.5 Effect of Beam Size

We show the effect of beam size in Figure 3.7. As beam size decreases, less of the power from the deflection angle is cut off (see Figure 3.4), and there is a significant improvement when going to higher resolution, up to about 1 arcmin beam (FWHM) for a $2 \times 10^{14} M_\odot$ cluster at $z = 0.7$. However, when the cluster is well resolved, the improvements become marginal. Smaller beam size may be useful in extracting other information from the cluster, such as concentration or other details about the profile.

3.4.6 Effect of Foreground Emission and other Secondary Anisotropies

Here we briefly comment on the impact of foregrounds on the matched filter estimator and we leave a full exploration with simulation and correlated emission to future work. We note that we expect many of the issues outlined in the detailed study of [158] in the context of a maximum likelihood approach to be relevant here as well, with some caveats that we now explain.

Firstly, the effects of residual foreground emissions uncorrelated with the cluster in question and with known power spectra can be incorporated in this formalism as an additional

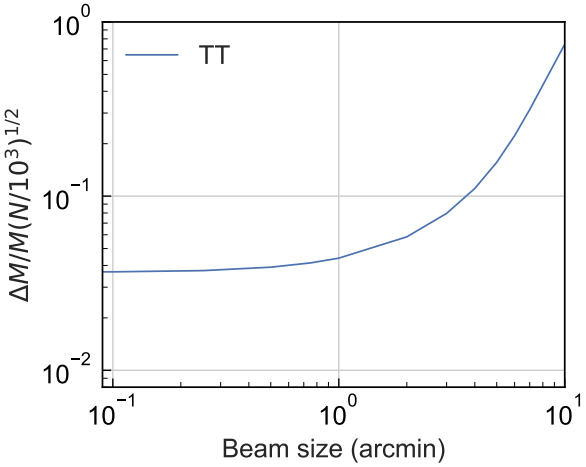


Figure 3.7: Effect of beam size on ability to extract mass, assuming 1 μK-arcmin instrumental noise, a $2 \times 10^{14} M_\odot$ cluster at $z = 0.7$, and a $N = 1000$ sample. At 0.3 arcmin effectively all the power of an NFW cluster is captured in the matched filter and no more improvements are possible by decreasing beam size.

source of noise contribution in Equation 3.25. In this case, the noise appearing in the plots would refer to "effective noise" after component separation, and we note that most of foregrounds can be in part mitigated by a multi-frequency analysis.

An exception is the kinematic Sunyaev-Zel'dovich (kSZ) effect caused by the bulk motion of free electrons ([190]). Unlike other foregrounds, the kSZ effect cannot be subtracted out by using multifrequency observations as it preserves the black body spectrum of the CMB, and therefore it represents an additional source of noise that is hard to overcome. As the measurements of the kSZ effect improve, it may be possible to construct a template of the emission and reduce the residual kSZ noise by a factor of order unity.

Any emission from the cluster itself, or from matter correlated with it should also only amount to extra noise, and not a bias, unless this emission is correlated with our estimate of the background gradient of the unlensed CMB. This is because the matched filter only detects dipole contributions aligned with the background gradient, due to the $\cos \phi_\ell$ term in Equation 3.29.

While not technically a foreground, gravitational lensing from uncorrelated structures in the universe produce extra small-scale anisotropy which again acts as a source of noise, which is included in the fiducial analysis throughout the paper. In Figure 3.6, we show that the difference in error caused by using the lensed vs unlensed CMB power spectra is only relevant at small noise levels. The effect of matter correlated with the halo is to enhance the lensing signal by what is usually known as the "two halo term". In a real analysis, the whole matter profile would be obtained and the one and two halo contributions would be fit

simultaneously, just as performed in most galaxy-galaxy lensing analyses.

3.5 Relation to previous work

Here we briefly compare the match filtered technique with other work in literature. Direct matched filtering techniques were first discussed as applicable to CMB cluster lensing in ([172]), and extensions in ([78, 203, 46]). Most of the subsequent work has focused on the use of quadratic estimators, as modified by ([84]) to ensure unbiased results ([136, 134]) This and other works based on the QE suffer from the statistical limitations discussed in our paper.=

Iterative techniques based on the quadratic estimator have been introduced in ([212, 213]). In the limit of large number of iterations the result should converge to the maximum likelihood solution in a unbiased way.

A maximum likelihood framework for cluster masses was developed in ([121, 15]) and further studied and extended to polarization in ([158]), which also provides extensive discussion of possible contamination from foregrounds. While it is not possible to outperform a full maximum likelihood approach, Figure 3.6, shows that for high redshift clusters or small noise limit, the matched filter presented in this paper approaches the maximum likelihood performance, but with only a small fraction of the computational cost.

On the experimental front, cluster lensing has first been detected using the quadratic estimator by ([128]) using data from the ACTPol experiment and CMASS galaxies from the BOSS survey. Subsequently, it has also been detected by the SPT collaboration using a Maximum-Likelihood technique ([15, 14]), as well as by the Planck collaboration ([2]).

Recently there has been work towards implementing a full maximum likelihood lensing map reconstruction without any assumptions about the shape of the lensing potential ([26]), based on previous work ([77]). Going forward, further developing these numerical techniques and applying them to small scale lenses would be very useful, but it currently appears challenging to obtain rapid convergence on small scales.

3.6 Summary

In this work we have explored the limitations of quadratic estimator reconstruction of CMB lensing from Temperature, in the small scale, low noise limit. We have shown that the quadratic estimator is fundamentally limited by suboptimal weighting, while in no-foregrounds, small noise limit, small lenses should be measurable arbitrarily well. This problem can be overcome by using different techniques, such as direct gradient inversion presented in this paper, or using more computationally expensive maximum likelihood methods.

An obvious application of the direct gradient inversion is reconstruction of cluster masses from CMB lensing. Once a profile for the potential is assumed, optimal weighting of the reconstructed modes is equivalent to a matched filter that we have explored in this work and

which out-performs the quadratic estimator over a wide range of instrumental sensitivities and redshifts; focused primarily on the low noise and small angular scale limit. The matched filtered approach relies on the antisymmetric nature of lensing as well as the separation of scales between cluster-lensing and the variation in the primary CMB. The antisymmetric nature of the matched filter is a useful feature which allows it to be less sensitive to kSZ and other foregrounds coming from the cluster.

The matched filter studied here can be extended to fit additional parameters such as the cluster concentration (see for example ([46])) or the amplitude of the two-halo term. Alternatively, tight priors can be put on them by studying the halo clustering, or detailed tSZ/X-ray observations. The effect of miscentering between the observed and true halo center needs to be quantified since it will lead to a bias if not properly accounted for, as pointed out by ([158]). At the same time, correlation between the selection function and other observables can produce biases in the inferred mean mass for the sample. For example, if halos elongated along the line of sight are preferentially selected, the inferred mass is likely to be overestimated due to the alignment of the halo with the surrounding cosmic structure. Detailed simulations and modeling are required to estimate the size of these effects for the particular sample being analyzed.

The technique discussed in this paper outperforms the QE for the noise level expected of future CMB experiments, and is computationally very efficient. Moreover, the noise calculation is analytically tractable, allowing direct forecasting and Fisher-type analysis without having to simulate a large number of clusters.

Acknowledgments

We appreciate helpful discussions with Eric Baxter, Anthony Challinor, Shirley Ho, Wayne Hu, Mathew Madhavacheril, Srinivasan Raghunathan, Emmanuel Schaan, Neelima Sehgal, Uro*s* Seljak, Kendrick Smith, and David Spergel. BH is supported by the NSF Graduate Research Fellowship, award number DGE 1106400. SF thanks the Miller Institute for Basic Research in Science at the University of California, Berkeley for support. BDS acknowledges support from an STFC Ernest Rutherford Fellowship.

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Chapter 4

Efficient Optimal Linear Reconstruction of Cosmological Density Field and Associated Bandpowers

4.1 Introduction

A ubiquitous problem of modern astrophysics is the reconstruction of the underlying signal from observed, noisy, and incomplete data. For linear fields the Wiener filter [211, 164] is the gold standard for reconstructing the underlying signal, as it is "optimal" in the sense that it minimizes the variance. It has been used as the basis of cosmological analysis for both large scale structure [60, 106] and CMB [21, 195].

However, the Wiener filter requires one to take the inverse of the overall covariance matrix, which has a noise and a signal based components. Noise is typically diagonal in observed space, while signal is diagonal in harmonic (or Fourier) space. In general, the covariance matrix will not be diagonalizable in either basis and it will be computationally difficult to invert the matrix numerically for a realistic sized survey map. While it is possible to make simplifying assumptions, like homogeneous and isotropic noise (as in, for example [109]), it is possible to efficiently implement Wiener filter by using the well studied property that the Wiener filter is mathematically equivalent to maximum a posteriori (MAP), which in turn can be solved with fast linear algebra methods or optimization [171].

This optimization can be performed using a variety of numerical techniques. Recent work used a messenger [53, 8] (or dual messenger [108]) field which can be diagonalized in either basis in which to run an approximation scheme. These methods have been argued to perform well versus other approximation schemes, but there are situations where the messenger field is zero (such as zero noise field with mask) and the method fails. In addition, it has been argued that a suitably chosen preconditioned conjugate gradient technique might be faster in

some instances [145]. Preconditioned conjugate gradient techniques have performed well in the context of CMB map reconstruction [96, 175], but require careful selection of the preconditioning scheme to achieve fast convergence. We therefore want to explore implementations that are both general and computationally efficient.

For cosmological analysis, it is not only the field that is of interest but also the bandpowers (such as power spectrum amplitudes) and their covariance matrix. Work towards estimating these quantities jointly with the underlying field has been done in the cosmic shear context [8, 7], but it required sampling from the joint probability distribution which is computationally involved. In this work we assume flat prior on band-powers and hence examine maximum likelihood estimation (MLE) method, after marginalization over the field, to quickly compute these quantities for observations with complex noise and mask properties as well.

In particular, we examine three cases of cosmological interest: linear density reconstruction, cosmic shear (E mode estimation alone as well as joint E/B), and primary CMB temperature anisotropy reconstruction. The outline of the paper is as follows: we discuss our MAP/MLE for field and power spectrum estimation in §4.2. In §4.3, we use these tools in a number of contexts, and compare the optimization approach with a numerically exact case in Subsection 4.3.2. In §4.4 we discuss our results and possible extensions of the work to analyze real data.

In Appendix A.1 we review the exact Weiner Filter approach with relies on inversion of the full covariance matrix. In Appendix 4.5 we discuss the convergence criteria in the case of primary CMB reconstruction. In Appendix 4.6 we discuss the joint estimation of E and B fields in the context of cosmic shear (although the technique transfers directly to CMB E and B polarization reconstruction as well).

4.2 Background

Here we summarize the optimization technique and standardize the notation. For a more through description, see [171, 180, 54, 174]. We measure quantities $d(\boldsymbol{r}_i)$ at select positions, such as a reconstructed projected density, forming a full data vector, $\boldsymbol{d}$. In general, this data vector will depend on a combination of underlying information about the field ("field coefficients") we wish to estimate, $\boldsymbol{s}$, and that which is pure noise, $\boldsymbol{n}$, i.e.

$$\boldsymbol{d} = \boldsymbol{R}\boldsymbol{s} + \boldsymbol{n}, \tag{4.1}$$

where the $\boldsymbol{R}$ is the response matrix expressing how our measurement changes with the underlying information. We express the underlying two point function information in terms of covariance matrices, $\boldsymbol{S} = \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle$, and $\boldsymbol{N} = \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle$, for the estimated signal and noise components, respectively. We assume that these are uncorrelated with each other, i.e. $\langle \boldsymbol{n}\boldsymbol{s}^\dagger \rangle = 0$; changes to this assumption are tractable but would require a redefinition of our underlying likelihood function and complicate the analysis since the noise would carry signal informa-

tion. The correlation matrix of the data is therefore,

$$\begin{aligned}\langle \boldsymbol{d}\boldsymbol{d}^\dagger \rangle \equiv C &= \langle (\boldsymbol{R}\boldsymbol{s}+\boldsymbol{n})(\boldsymbol{R}\boldsymbol{s}+\boldsymbol{n})^\dagger \rangle \\ &= \langle (\boldsymbol{R}\boldsymbol{s}(\boldsymbol{R}\boldsymbol{s})^\dagger + \boldsymbol{n}\boldsymbol{n}^\dagger + \text{Cross Terms} \rangle \\ &= \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^\dagger + \boldsymbol{N}. \end{aligned} \tag{4.2}$$

Our signal covariance, $\boldsymbol{S}$, takes a diagonal form in Fourier space. The noise covariance, $\boldsymbol{N}$, is often approximately diagonal in data space, as there usually are no noise correlations between data. We can convert the covariance matrix into Fourier basis where signal covariance is diagonal, but this will lead to very strong off-diagonal terms of the noise matrix, in the presence of masked sky regions and/or variable noise properties. Therefore, our overall data correlation matrix cannot be diagonalized easily in either basis.

We can re-express our covariance in terms of underlying band-powers, $\boldsymbol{\Theta}$, labeling each modes of interest to estimate as $\{1, \cdots, l, \cdots, l_{max}\}$, and the rest as $\{l_{max+1}, \cdots, \infty\}$, and then expressing our correlation matrix as

$$\boldsymbol{C} = \sum_l \Theta_l \boldsymbol{Q}_l + \boldsymbol{N}. \tag{4.3}$$

This new $\boldsymbol{Q}_l = \Theta_l \boldsymbol{\Pi}_l \boldsymbol{R}\boldsymbol{R}^\dagger \boldsymbol{\Pi}_l$ basis is the projection (with projection operators $\boldsymbol{\Pi}_l$) of the response matrix $\boldsymbol{R}$ for each mode. The band-power can correspond to averaging over spherical harmonic basis coefficients at a constant l, adding up $2l+1$ m modes, but we can also average over more than one l.

4.2.1 Bandpowers posterior

We now want to find the most probable set of bandpowers for a given set of measurements $\boldsymbol{d}$, assuming flat prior on band-powers. We thus parametrize the power spectrum as a function of these bandpowers $\boldsymbol{S}(\boldsymbol{\Theta})$. If we assume our modes are Gaussian we can express the likelihood function in the familiar form, i.e.

$$L(\boldsymbol{d}|\boldsymbol{\Theta}) = (2\pi)^{-N/2} \det(\boldsymbol{C})^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{d}^\dagger \boldsymbol{C}^{-1} \boldsymbol{d}\right). \tag{4.4}$$

Associated with the likelihood function and a parameter set $\hat{\boldsymbol{\Theta}}$ which maximizes it, is the Hessian matrix [**BondJaffeEtAl98**],

$$F_{ll'} = -\frac{\partial^2 \ln L}{\partial \Theta_l \partial \Theta_{l'}}. \tag{4.5}$$

The inverse of the Hessian matrix can be interpreted as an estimate of the covariance matrix of the parameters, i.e.

$$\boldsymbol{F}^{-1} = \langle \boldsymbol{\Theta}\boldsymbol{\Theta}^\dagger \rangle - \langle \boldsymbol{\Theta} \rangle \langle \boldsymbol{\Theta} \rangle^\dagger. \tag{4.6}$$

We now have the bandpower posterior in the gaussian form, given by the mean $\hat{\boldsymbol{\Theta}}$ and the covariance matrix $\boldsymbol{F}^{-1}$. To obtain the solution for the mean it is easiest to use Newton's second order method, which gives a quadratic estimator of the form [194]

$$\Theta_l = \frac{1}{2}\sum_{l'} F^{-1}_{ll'}(\boldsymbol{d}^\dagger \boldsymbol{C}^{-1}\boldsymbol{Q}_{l'}\boldsymbol{C}^{-1}\boldsymbol{d} - b_{l'}), \tag{4.7}$$

where b_l is a noise bias term that can be found by computing the ensemble average of the first term assuming $\theta_l = 0$ for all modes probed (i.e. $l < l_{max}$),

$$b_l = \mathrm{tr}\left[\boldsymbol{N} + \sum_{l_{max}+1}^{\infty}(\Theta_l \boldsymbol{Q}_l)\boldsymbol{C}^{-1}\boldsymbol{Q}_l\boldsymbol{C}^{-1}\right]. \tag{4.8}$$

4.2.2 MAP Field Reconstruction

In practice this analytical calculation requires the inversion of a large matrix, $\boldsymbol{C}$, which does not necessarily have properties that make inversion efficient (i.e. block diagonal or sparse) and will in general require $O(n^3)$ time for an $n \times n$ matrix. In the case of reconstructing the underlying density field for astronomical large surveys with n pixels, this would be prohibitively computationally expensive for the foreseeable future. Instead, we will approach this as an optimization problem [174]. We will not use 4.4, and instead of expressing the likelihood of data given bandpowers we will work in terms of latent variables, writing the joint distribution of $\boldsymbol{s}$ and $\boldsymbol{d}$,

$$p(\boldsymbol{s},\boldsymbol{d}|\boldsymbol{S}) = (2\pi)^{-(N+M)/2} det(\boldsymbol{S}\boldsymbol{N})^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{s}^\dagger\boldsymbol{S}^{-1}\boldsymbol{s} + (\boldsymbol{d}-\boldsymbol{R}\boldsymbol{s})^\dagger\boldsymbol{N}^{-1}(\boldsymbol{d}-\boldsymbol{R}\boldsymbol{s})\right), \tag{4.9}$$

and note that the minimum variance solution for the modes can be found by minimizing the loss function χ^2,

$$\chi^2 = -2\ln p(\boldsymbol{s},\boldsymbol{d}|\boldsymbol{S}) + c = \boldsymbol{s}^\dagger\boldsymbol{S}^{-1}\boldsymbol{s} + (\boldsymbol{d}-\boldsymbol{R}\boldsymbol{s})^\dagger\boldsymbol{N}^{-1}(\boldsymbol{d}-\boldsymbol{R}\boldsymbol{s}), \tag{4.10}$$

with respect to $\boldsymbol{s}$. At m-th iteration we are at $\boldsymbol{s_m}$, where we have

$$\chi^2 = \chi^2_0 + 2\boldsymbol{g}(\boldsymbol{s}-\boldsymbol{s_m}) + (\boldsymbol{s}-\boldsymbol{s_m})^\dagger\boldsymbol{D}(\boldsymbol{s}-\boldsymbol{s_m}), \tag{4.11}$$

with gradient function in terms of the derivative of the response function, $\boldsymbol{R}$, given as

$$\boldsymbol{g} = \frac{1}{2}\frac{\partial\chi^2}{\partial\boldsymbol{s}} = \boldsymbol{S}^{-1}\boldsymbol{s_m} - \boldsymbol{R}^\dagger\boldsymbol{N}^{-1}(\boldsymbol{d}-\boldsymbol{R}\boldsymbol{s_m}). \tag{4.12}$$

For the linear problems studied in this work, this derivative can be calculated analytically, but in other more involved cases (such as nonlinear structure formation) might be computationally involved as it would require intensive back-propagation. The solution where

$\boldsymbol{g} = 0$, and therefore a local extremum is found, will be denoted $\hat{\boldsymbol{s}}$, and is the maximum a posteriori solution (MAP). For linear problems it is the best possible solution in the sense to minimizing the variance.

The curvature matrix $\boldsymbol{D}$ has the form

$$\boldsymbol{D} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \boldsymbol{s} \partial \boldsymbol{s}} = \boldsymbol{S}^{-1} + \boldsymbol{R}^\dagger \boldsymbol{N}^{-1} \boldsymbol{R}. \tag{4.13}$$

However, in this work we will not explicitly evaluate it, as it is too large. Instead, we will use low rank approximation as performed by L-BFGS quasi-Newton optimization method. We will use L-BFGS as the optimization method in this paper.

The starting point for the optimization algorithm does not play a significant role for linear problems as the posterior surface is convex and the true global minimum can always be found. In practice, for the examples in this work, we found no noticeable effects of the starting point in terms of convergence properties, i.e. required number of iterations.

4.2.3 Minimum Variance Estimation of the Power Spectrum

The result of the above optimization procedure is $\hat{\boldsymbol{s}}$, and is useful for creating maps, but has more information than needed for cosmological analysis. If our goal is to determine a set of summary statistics/band-power measurement, $\boldsymbol{\Theta}$, such as a power spectrum bandpowers, we need to marginalize over the latent variables, the modes $\boldsymbol{s}$. To do so we need to define a projection matrix $\boldsymbol{\Pi}_l$ around a fiducial power-spectrum $\boldsymbol{S}_{\rm fid}$ with associated band-powers $\Theta_{\rm fid}$, defined as

$$\left[\frac{\partial \boldsymbol{S}}{\partial \Theta_l}\right]_{\boldsymbol{S}_{\rm fid}} = \boldsymbol{\Pi}_l. \tag{4.14}$$

This fiducial power spectrum is a regularized version of the measured power spectrum, and is thus iterated upon: we start with some fiducial prior, which we then update if the data require us to do so. This process is regularized, i.e. we use a smooth version of the measured power spectrum, for example a power spectrum predicted by the cosmological parameters we are determining from these data.

The true covariance can be written in terms of this fiducial power-spectrum plus a term linear in the projection operators:

$$\boldsymbol{S} = \boldsymbol{S}_{\rm fid} + \sum_l \Delta\Theta_l \boldsymbol{\Pi}_l, \tag{4.15}$$

where $\Delta\Theta_l$ is the difference of the band-powers to those of the fiducial model. For the cases studied in this work, the dependence of $\boldsymbol{S}$ on $\boldsymbol{\Theta}$ is linear so we can take

$$\boldsymbol{\Pi}_l = \frac{\boldsymbol{S}_{\rm fid}}{\Theta_l}, \tag{4.16}$$

i.e. the projection matrix takes the power spectrum per bin, $\boldsymbol{\Theta}_l$, to the full power spectrum, $\boldsymbol{S}$. Note that the choice of the fiducial model is important in that if it is sufficiently far

away from the true model the result could be biased, but iteratively recalculating $\boldsymbol{S}_{\mathrm{fid}}$ with the solved new band-powers adjusted by $\Delta\boldsymbol{\Theta}$ will provide an asymptotically more accurate reconstruction. In the cases of interest in this work, no iterative process was necessary to provide an accurate reconstruction. In practice for examples in this work, we used the true power-spectrum with each power re-scaled by a random $O(0.1)$ value.

We are assuming flat prior for the bandpowers, so to compute the posterior distribution of band-powers we can write their (marginalized over $\boldsymbol{s}$) likelihood function to maximize as a second order expansion around the fiducial model

$$\ln L(\boldsymbol{\Theta}_{\mathrm{fid}}+\Delta\boldsymbol{\Theta}) = \ln L(\boldsymbol{\Theta}_{\mathrm{fid}}) + \sum_l \left[\frac{\partial \ln L(\boldsymbol{\Theta})}{\partial\Theta_l}\right]_{\boldsymbol{\Theta}_{\mathrm{fid}}} \Delta\Theta_l + \frac{1}{2}\sum_{ll'}\left[\frac{\partial^2 \ln L(\boldsymbol{\Theta})}{\partial\Theta_l\partial\Theta_{l'}}\right]_{\boldsymbol{\Theta}_{\mathrm{fid}}} \Delta\Theta_l\Delta\Theta_{l'}; \tag{4.17}$$

where we assume a flat prior on the band-powers.

We define

$$E_l(\boldsymbol{S}_{\mathrm{fid}},\hat{\boldsymbol{s}}) = \frac{1}{2}\hat{\boldsymbol{s}}^\dagger \boldsymbol{S}_{\mathrm{fid}}^{-1}\boldsymbol{\Pi}_l\boldsymbol{S}_{\mathrm{fid}}^{-1}\hat{\boldsymbol{s}} = \frac{1}{2}\sum_{k_l}\frac{\hat{s}_{k_l}^2}{\Theta_{\boldsymbol{fid},l}S_{\mathrm{fid},k_l}}, \tag{4.18}$$

where in the last expression we define the sum over k_l as the sum over all modes which contribute to band-power, Θ_l, and in the last equality we made use of the diagonal property of the projection operators and fiducial power spectrum. Putting this together we find that the derivative of the likelihood can be expressed as [174]

$$\frac{\partial \ln L(\boldsymbol{\Theta})}{\partial\Theta_l} = E_l - b_l, \tag{4.19}$$

where we defined

$$b_l = \frac{1}{2}\mathrm{t}r\left[\frac{\partial \ln(\boldsymbol{SN})}{\partial\Theta_l}\right]_{\boldsymbol{S}_{\mathrm{fid}}}. \tag{4.20}$$

For the linear cases studied in this work, this term is often called the noise bias term. However, it is worth remembering that this term's origin is the derivative of the log determinant of the product of the Hessian and the signal covariance matrices in equation 4.20 (since noise covariance derivative is zero). To find MLE we need to find the zero of Eq. 4.19, which we solve using Newton's method. To do this we define the Hessian matrix,

$$F_{ll'} = -\frac{\partial^2 \ln L(\boldsymbol{\Theta})}{\partial\Theta_l\partial\Theta_{l'}}, \tag{4.21}$$

which for linear models defines the Gaussian posterior assuming sufficient modes have been averaged over so that by central limit theorem we can describe the posterior as a multivariate gaussian. The peak of the likelihood function can be found by setting the derivative of equation 4.17 with respect to $\boldsymbol{\Delta\Theta}$ to zero, which upon inserting equation 4.19 yields

$$(\boldsymbol{F}\Delta\hat{\boldsymbol{\Theta}})_l = E_l - b_l. \tag{4.22}$$

4.2.3.1 Estimation of the Noise Bias and the Hessian

While the noise bias, b_l, and the Hessian matrix, $\boldsymbol{F}$, from Equation 4.22 could be calculated exactly, this will involve inversion of large matrices, which is what we are trying to avoid by deriving the MAP via optimization techniques. Instead, we will perform a simulation based analysis motivated by the underlying definition of each of these terms.

In general, the maximum likelihood field, $\hat{\boldsymbol{s}}$, attained with the procedure described in Sec 4.2.2, will have bias due to the presence of noise: when the noise is high the minimum variance estimator drives $\boldsymbol{s}$ to zero. In the case of cosmological density fields which have red power spectra (less power on small scales compared to white noise), this will result in washing out the small scale power. See the figures in Sec 4.3 for explicit examples.

To correct for this bias we need to understand how our reconstruction responds to the presence of noise. For this we perform a simulation analysis wherein we generate a data realization generated from a fiducial power spectrum, inject the noise and mask, perform the optimization and see how the presence of noise affects the reconstruction. Let us call the new data and noise realization data $\boldsymbol{d}_{s+n}$, with associated maximum likelihood reconstruction $\hat{s}_{s+n}$. The gradient of equation 4.19 has to vanish if evaluated at the fiducial model. The noise bias in this case can be found directly as

$$b_l = E_l(\boldsymbol{\Theta}_{\rm fid}, \hat{\boldsymbol{s}}_{s+n}). \tag{4.23}$$

This quantity should be averaged over many realizations, but for the linear signal-dominated cases studied in this work we found even one realization was sufficient for an accurate reconstruction.

To calculate the Hessian matrix, we evaluate the gradient of equation 4.19 at two different fiducial model values, and use finite differentiation [174],

$$F_{ll'}\Delta\Theta_{l'} = E_l(\boldsymbol{\Theta}_{\rm fid} + \Delta\Theta_{l'}) - E_l(\boldsymbol{\Theta}_{\rm fid}). \tag{4.24}$$

Its inverse is the covariance matrix for the band-powers. This is in contrast to directly using linear algebra techniques to calculate the Hessian matrix (see Equation A.6) which would be numerically intractable for a realistic survey size. Using Equation 4.24 in the linear case, one can calculate it at the cost of additional optimization step. Since this is a linear problem Hessian matrix equals Fisher information matrix and thus gives the smallest attainable errors on the parameters (Cramér-Rao theorem).

4.2.4 Procedure Summary

1. Initialize a Gaussian random field (the true signal field) with some underlying power-spectrum.

2. Apply the response operator to this field, and additional noise and masking terms. The output of this is the input data vector, $\boldsymbol{d}$.

3. An estimate of the underlying signal field is created through optimization as described in Subsection 4.2.2 to yield $\hat{\boldsymbol{s}}$.

4. An initial estimate of the band-powers is generated by taking the power spectra of the reconstructed map and binning.

5. To this band-power estimate, we apply the noise bias correction (estimated using Eq 4.23) and Hessian matrix (estimated using Eq 4.24) to provide an optimal reconstructed value of the band-powers given in Eq 4.22.

4.3 Example Cases

Here we implement the above scheme in a number of simulated cosmological contexts to demonstrate its versatility and efficiency. For these cases, we set our convergence criteria to be $\epsilon \equiv \delta\chi^2 = 10^{-1}$; i.e. the optimization ends when the difference of the absolute chi-square values between iterations is 10^{-1} (typical value of χ^2 is of order 2×10^5 for the dimensionality used here). In Appendix 4.5 we discuss the choice of this criteria in the context of CMB reconstruction, but we have found it to be sufficient for all the example cases.

4.3.1 Projected Density Field

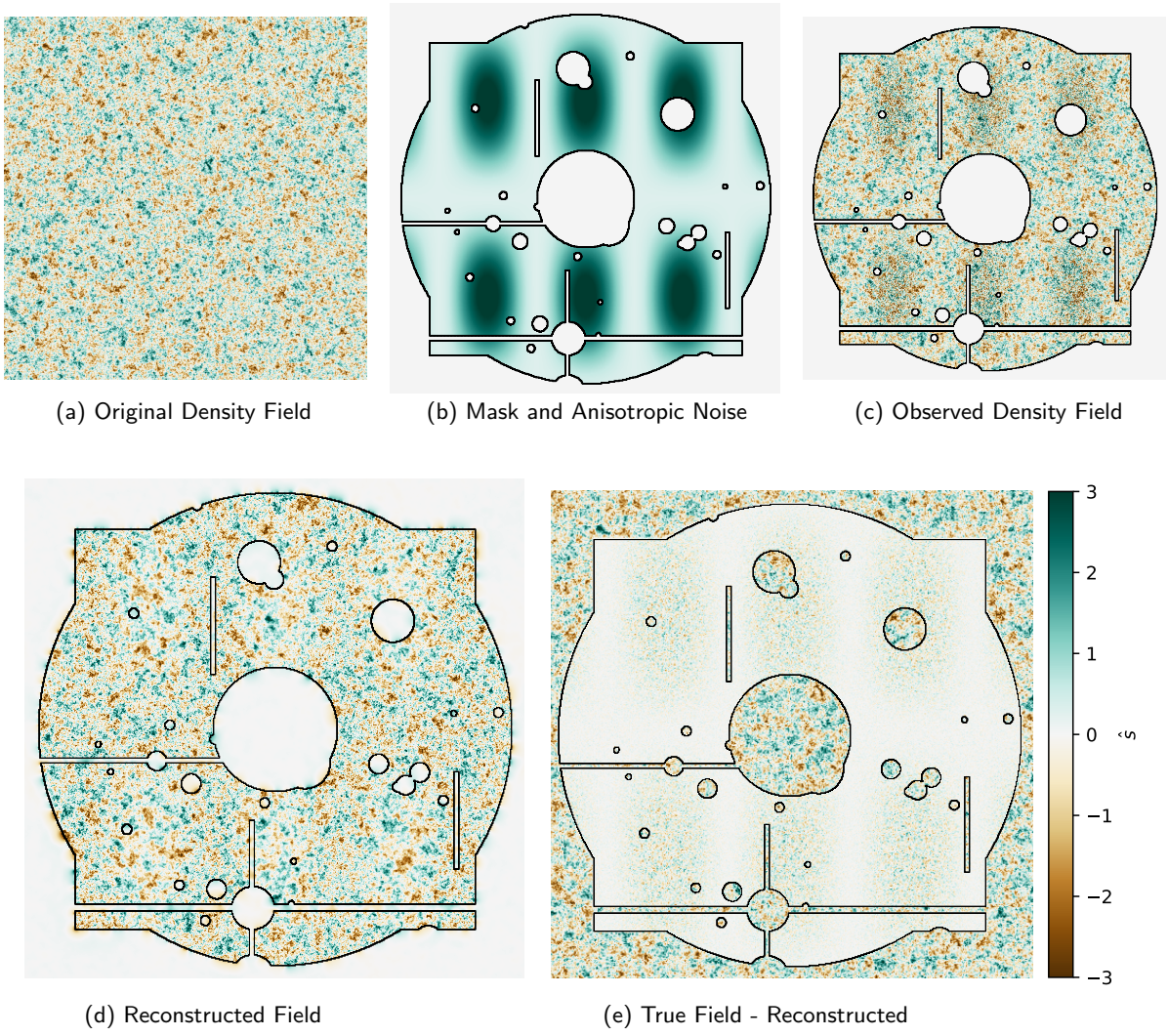
As our first case, we will look at reconstructing the density field from a noisy measurement of the density field. The starting measurement could come from a variety of observations such as galaxy clustering or Lyman alpha forest tomography. For a complete analysis one would specify the response matrix $\boldsymbol{R}$ to go from the density to the observables which would include a model for the various biases present in the observations. In this case we assume the bias model is known and deal directly with the matter density field.

We generate a Gaussian random field with a power-spectrum given by

$$P(k) \propto \frac{k}{1 + k^3} \tag{4.25}$$

over a 2D, $L = 1380$ Mpc/h side-length box. We introduce an anisotropic white noise over the field to simulate either irregularities in depth of a given survey or theoretical uncertainties in the underlying bias model. In Appendix A.1 we demonstrate the validity of the L-BFGS optimization method in a small test case (64×64 pixel) where it is also tractable to exactly invert the full covariance matrix numerically thereby providing validation of our maximum likelihood technique, while here we examine a more realistic 512×512 pixel map. We also use a realistic mask which includes foreground stars and other potential image defects.

Using the input power spectrum of Eq. 4.25, we generate a density field shown in Fig 4.1(a), apply a mask and anisotropic noise shown in Fig 4.1(b), which results in a mock observation in Fig 4.1(c). We perform the minimization routine outlined in Sec. 4.2.2, with the optimized map shown in Fig. 4.1(d) with residuals shown in Fig 4.1(e). Qualitatively the



(a) Original Density Field (b) Mask and Anisotropic Noise (c) Observed Density Field

(d) Reconstructed Field (e) True Field - Reconstructed

Figure 4.1: MAP density reconstruction for the 512x512 pixel case. Note that images (a), (c), (d),and (e) have same absolute color scale, while (b) shows the spatial variance of the noise properties. Color scale is normalized to show standard deviations away from mean.

field is accurately reconstructed within the mask in the low-noise regions and is even able to reconstruct the larger scale modes right on the border within the masked region. However, as expected, the small scale modes within the high noise regions within the mask are poorly reconstructed since it is impossible to differentiate those modes in real space with the noise. In addition, small masked regions have very low residual error as there are sufficient, well sampled, nearby large scale modes to infer the regions value.

In Fig 4.2 we show the comparison of the optimized result with the true power-spectra

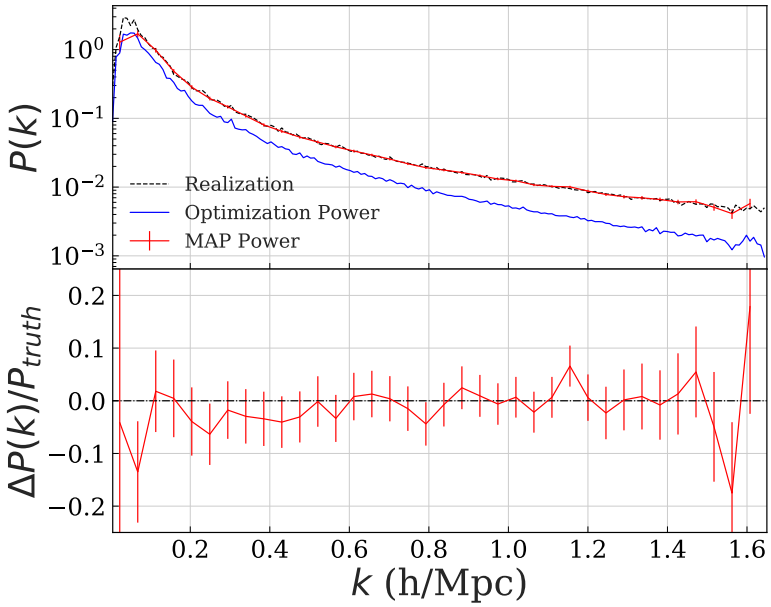


Figure 4.2: Comparison of the maximum likelihood power spectrum attained from optimization versus the true power-spectrum of the region for the density field. Also shown is the importance of the noise bias correction (or, equivalently, the importance of the Hessian determinant).

of the entire field. We also show the effect of the noise bias correction, which in this case is substantial as small scale power is washed out in the high noise regions as well as due to the masked regions. However, this power is recoverable using the analysis described in Section 4.2.

In Fig 4.3 we show how the number of iterations of the reconstruction algorithm scales with the box size. Each iteration requires a calculation of the loss function and the derivatives of the field which can be found analytically. Notice that the iteration number increases with the box size, but only a factor of a few when going from 64^2 to 1024^2. This does not include estimation of the noise bias and the Hessian matrix, which will depend on the number of bandpowers. As each row of the Hessian matrix requires an additional optimization, the true number of iteration will scale linearly with the number of bandpowers. In practice, since the Hessian matrix is very smooth and nearly translationally invariant (in this case, it is peaked on the diagonal and monotonically decreasing away from the diagonal), one simply needs to sample the matrix along a small number of rows and interpolate between them.

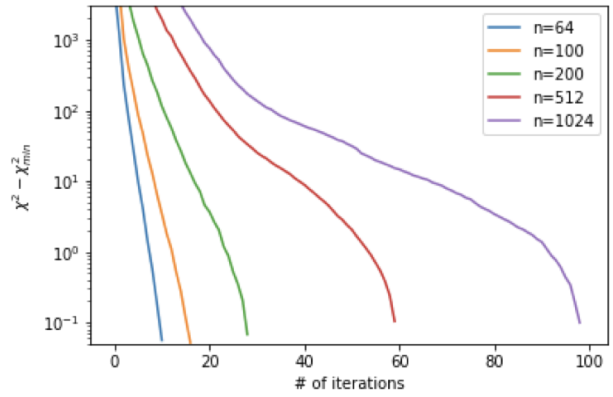


Figure 4.3: Convergence properties of the L-BFGS estimator as a function of the box-side dimension n.

4.3.2 Wiener Filtering vs. MAP Projected Density Example

Direct numerical evaluation of Wiener filtering is computationally expensive as it requires the direct inversion of a matrix with the square of the number of pixels in the survey (see App A.1), so we specialize our direct comparison to a small 64×64 pixel image.

Using the input power spectrum of Eq. 4.25 we generate a density field shown in Fig 4.4(a), apply a mask and anisotropic noise shown in Fig 4.4(b), which results in a mock observation in Fig 4.4(c). We perform the minimization routine outlined in Sec. 4.2.2, with the optimized map shown in Fig. 4.4(d) with residuals shown in Fig 4.4(e). There is qualitative agreement between the truth and the reconstructed field within the mask in the low-noise regions and is even able to reconstruct the larger scale modes right on the border within the masked region. However, as expected, the small scale modes within the high noise regions within the mask are poorly reconstructed since it is impossible to differentiate those modes in real space with the noise.

For this small test case we can compare the optimization result directly against a numerical inversion Wiener filter solution, which we show on the map level in Fig 4.5. The results match outside the masked region within 10^{-4}, while in the masked region there is a slightly large difference due to the imposition of a convergence criteria in our L-BFGS scheme. As we increase the required precision of the L-BFGS in terms of $\epsilon \equiv \Delta\chi^2$, we asymptotically approach the Wiener filter solution.

Using the formulation in Sec 4.2.3, we can look at the performance of the technique as a function of scale. In Fourier space we can account for the reduction of small scale power caused by noise and also estimate the Hessian matrix (thereby giving error estimates). We show the power spectrum and error estimates from the optimization technique versus the direct Wiener filtering in Fig 4.6. Note that the full reconstruction relies on both calculation

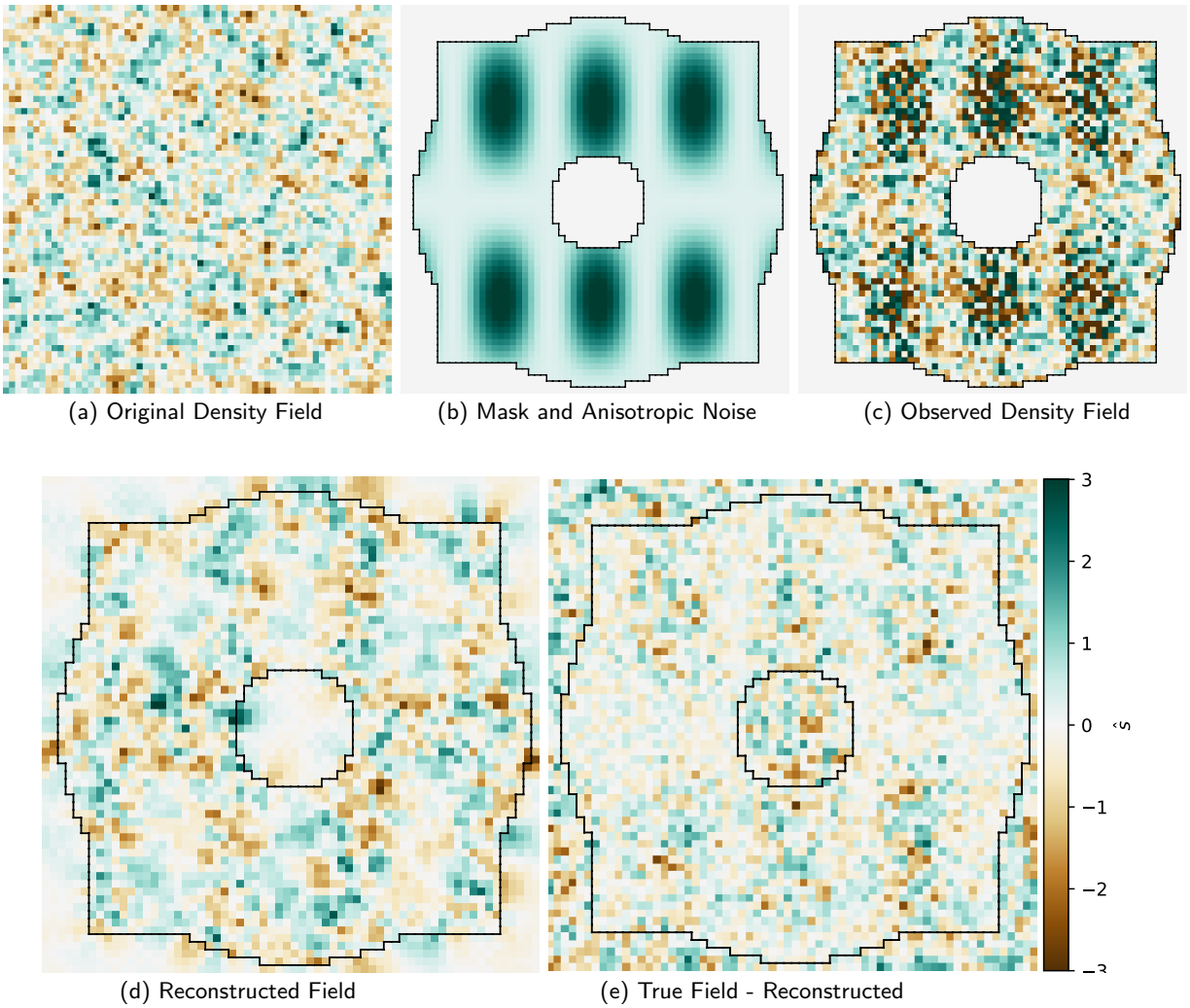


Figure 4.4: MAP density reconstruction for the 64x64 case.

of noise bias and Hessian matrix. We have compared each of these terms from the optimization method to those calculated via direct matrix inversion to confirm they are equal within the error of the required optimization precision. Also note that we only used one noise realization to estimate the noise bias. In general, the number of noise realizations necessary to appropriately estimate the number of underlying band-powers will depend on both the underlying noise model and the band-powers of interest. In this particular case we found the improvements from including multiple noise realizations minimal as the effect on the overall power-spectrum were sub 1%.

We also explored the noise dominated regime more explicitly in Figure 4.7, where we apply a uniform high noise level over the entire field with variance 1.5 times the average

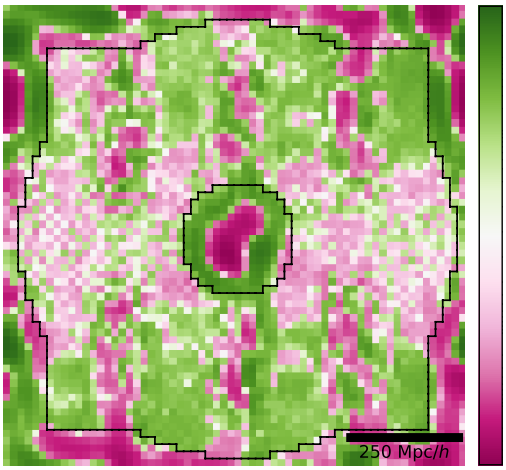


Figure 4.5: The log absolute magnitude difference of the direct matrix inversion Wiener filter solution and that attained via a MAP method. Note that differences are extremely small throughout the map and are particularly small in the unmasked region.

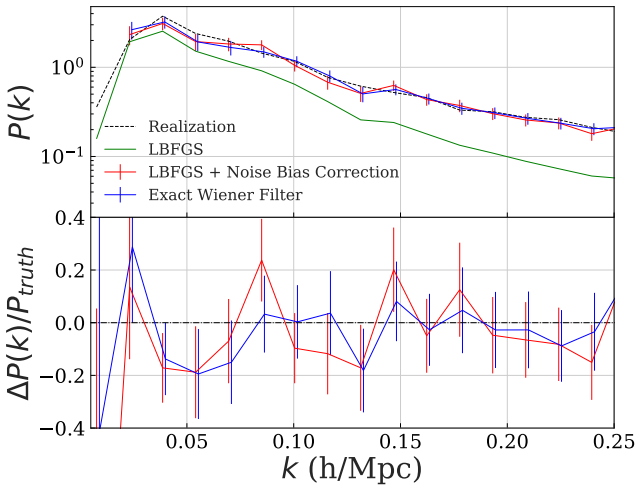
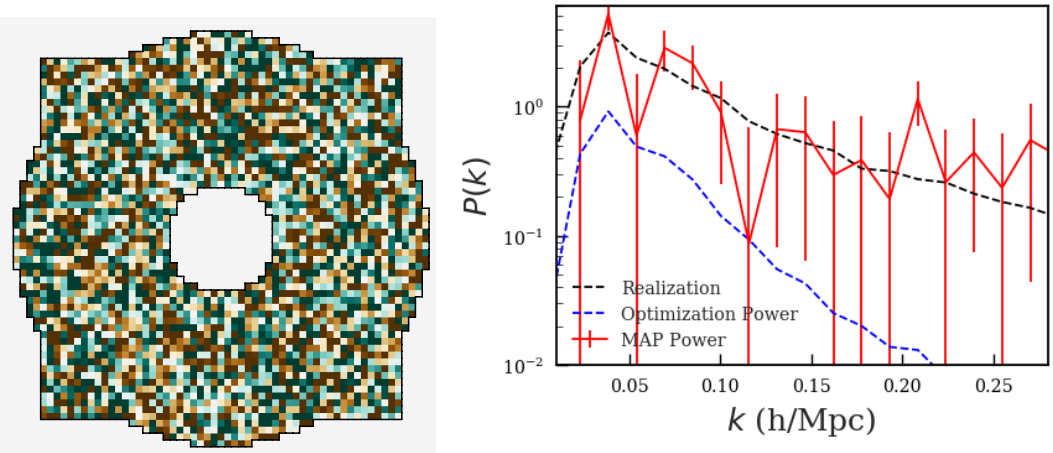


Figure 4.6: Comparison of the maximum likelihood power spectrum attained from optimization versus that for brute force matrix inversion Wiener filter. Errors are visually identical, and all points k bins are within one standard deviation of each other.



(a) High Noise Density Field (b) Reconstructed Bandpowers

Figure 4.7: Observed field and associated reconstructed bandpowers for the noise dominated regime. Mask and color scale are the same as in Fig 4.5, and noise is uniform over the field.

variance in density. In this regime the optimized power is significantly suppressed across all scales, but the modes are still recoverable.

4.3.3 Cosmic Microwave Background Temperature

The question of optimal reconstruction of CMB maps given irregular sky coverage and variable noise and foreground subtraction is a common issue for existing CMB surveys. So far, true maximum likelihood power spectra estimators have only been applied to data from WMAP [195] and for the largest angular scales in Planck data [152], but these techniques are difficult to scale to the entire Planck dataset due to the significant increase in computational cost.

While Planck's power spectrum measurements, and therefore cosmological parameter constraints [153], do not rely on construction of the actual full map, other spatially dependent signals do. Cross-correlations between the primary CMB and other cosmological probes, such as x-ray signal or galaxy positions, require an accurate spatial reconstruction of the CMB map. In addition, full sky CMB lensing maps are constructed by applying the quadratic estimators to the CMB map (in either temperature, polarization, or some minimum variance combination of the two) and will similarly suffer if the reconstructed temperature and/or polarization maps are suboptimal. In terms of map reconstruction, there are highly efficient Wiener filter programs available [175] which perform a multi-scale analysis with preconditioned optimization routines, which could replace our L-BFGS, but since the number of iterations needed is already low we did not explore this further.

In Figure 4.8 we show results for temperature, without polarization. We mention that

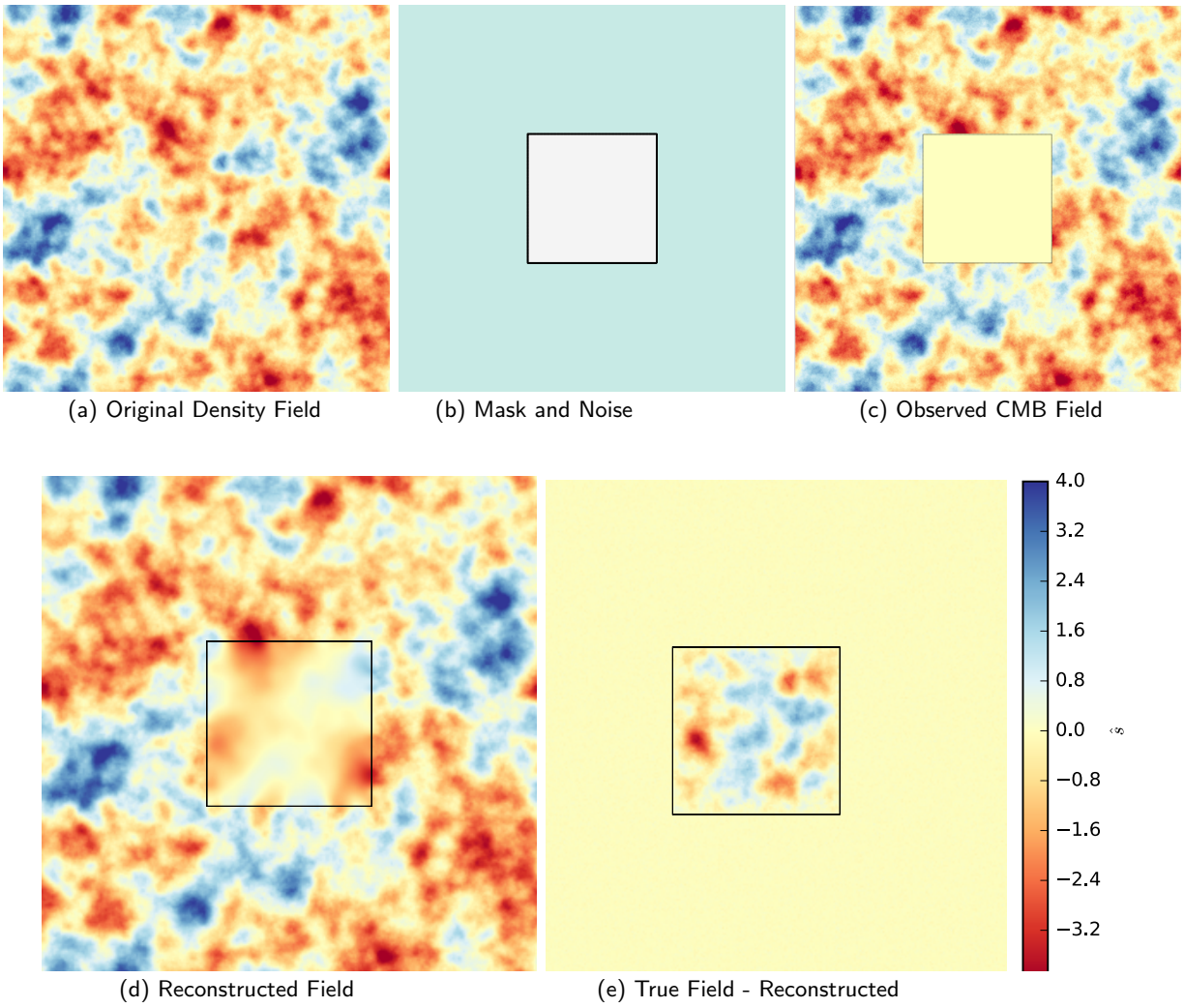


Figure 4.8: MAP CMB reconstruction for 512x512 pixel map. Note that images (a), (c), (d),and (e) have same absolute color scale, while (b) shows the spatial variance of the noise properties. Color scale is normalized to show standard deviations away from mean.

polarization would be analogous to the example of joint optimization of E and B fields, which we do in the case of cosmic shear in Appendix 4.6.

The case of CMB reconstruction is analogous to that of the density field, but with a significantly redder spectrum. The condition number of the covariance matrix is thus significantly larger, and BFGS needs more iterations to converge. While we could have used conjugate gradient with a preconditioner (as in [175]) to improve the convergence of the optimization step, we chose not to do so here since the computational cost was not significantly higher than that for the simple density case even for this case (see Figure 4.11). For the

implementation of our algorithm it is important to recognize that the increase in power on larger scales in the CMB case makes the masked region the most computationally expensive region to reconstruct, and a redder spectrum will allow more mode reconstruction within this region. For more discussion, see Appendix 4.5.

We generate a mock primary CMB full-sky field using HEALPIX [65] based on power spectrum generated from CLASS using the Planck 2015 cosmological parameters [153]. We then extract a 10×10 degree patch which we then mask a central region and introduce a white noise of 6 μK-arcmin.

The reconstruction is quite good in the observed region as the presence of noise mostly affects small scales where there is very little power. In the masked region the optimization is able to reproduce some clear long-scale modes. In Appendix 4.5 we show that with a more accurate convergence criteria we reproduce more of these large scale modes in the masked region but in terms of power spectrum estimation this is unnecessary, as we are already cosmic variance limited even with $\epsilon = 0.1$.

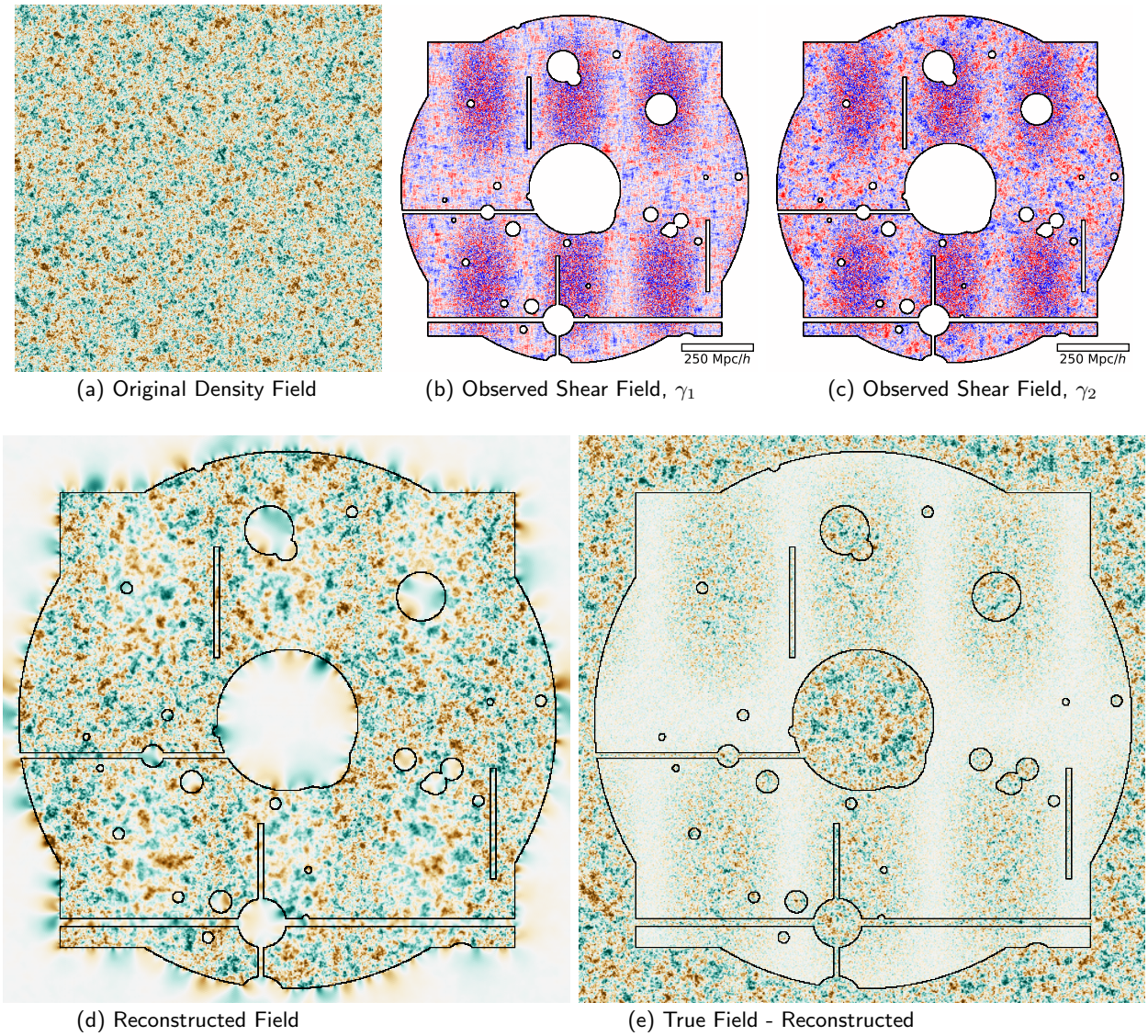
4.3.4 Cosmic Shear

In this section we specialize to only fitting a curl-free E component; we discuss the more general case including a curl component in Appendix 4.6. For a details on the cosmic shear formalism, see [204, 13, 91].

To apply our method, we perform optimization over the underlying density field and at each step of the optimization compute the corresponding shear maps, γ_1 and γ_2, to compare with the mock observed shear maps. In principle, instead of working with the shear maps which require binning and/or interpolation between galaxies, one could work directly with the catalog of galaxy shapes and compute the likelihood of the observed ellipticity instead of the given averaged shear maps (as in [143, 17]).

This method is in contrast to the standard Kaiser-Squires (KS) [99] technique which has proven quite successful so far in cosmic shear analysis and cluster mass estimation. However, KS has some notable downsides in the presence of anisotropic noise or a mask as it is not able to self consistently down-weight the high noise areas and masked regions, resulting in defects on boundaries. In addition, the noise inherent in these measurements propagates onto the final mass-maps, resulting in an inaccurate small scale power measurement. This has been shown to be particularly detrimental to peak statistics measurements [95].

The results in position space are shown in Figure 4.9. In Figure 4.9(a) we show the initial density field and in Figure 4.9(b,c) we show the observed shear fields including mask and noise properties. Our reconstructed maximum likelihood map is shown in Figure 4.9(d) and the difference between the original field and reconstructed in Figure 4.9(e). As in the case of the density field, the optimization technique is able to exactly reconstruct the density in the low noise, unmasked regions, but only recovers large scale scale power in the higher noise unmasked regions.



(a) Original Density Field (b) Observed Shear Field, γ_1 (c) Observed Shear Field, γ_2

(d) Reconstructed Field (e) True Field - Reconstructed

Figure 4.9: Maximum likelihood shear reconstruction for a 512x512 pixel map. Here we use the same mask and noise properties, as well as color scaling, as shown in Figure 4.1.

4.4 Discussion and Conclusion

In this work we have demonstrated that is is possible to efficiently reconstruct the MAP signal field and the maximum likelihood power spectrum for linear fields for realistic survey sizes. This technique is equivalent to the Wiener filter solution for small enough convergence criteria and has been applied to a number of cosmological fields (density, CMB, and cosmic shear maps). We are able to reconstruct the initial density field and the overall power spec-

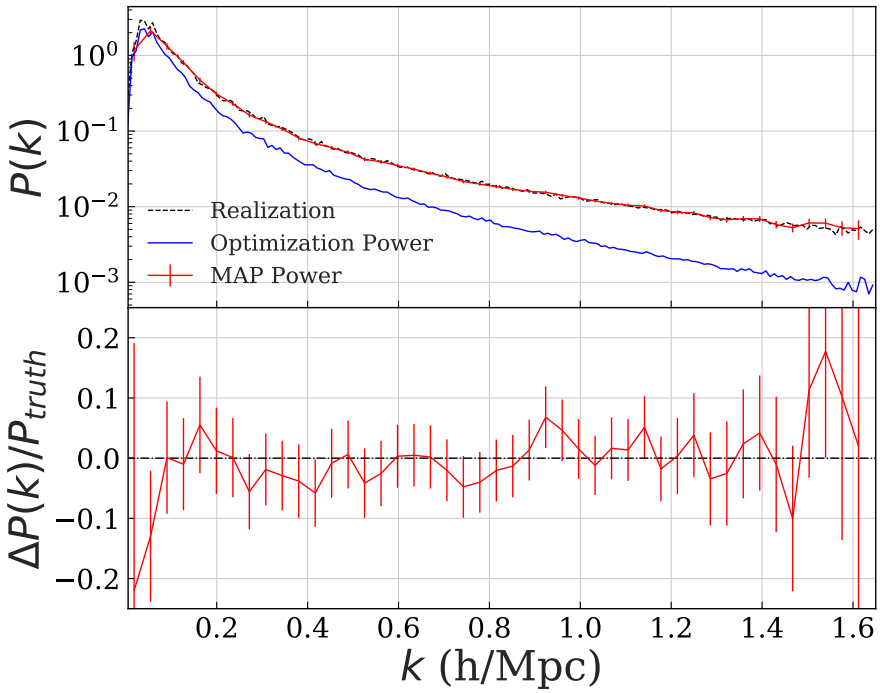


Figure 4.10: Comparison of the maximum likelihood power spectrum attained from optimization versus the true power-spectrum of the region for the shear-only reconstruction. Also shown is the importance of the noise bias correction (or, equivalently, the importance of the Hessian determinant).

trum, accounting for noise bias and window-function effects due to masking. We first recast the field reconstruction into an optimization problem, which we solve using quasi-Newton optimization. We then recast the power spectrum estimation into the field marginalization problem, from which we obtain an expression that depends on the field reconstruction solution and a determinant term. We develop a novel simulation based method for the latter. We extend the simulations formalism to provide the covariance matrix for the power spectrum.

This technique outperforms the brute force Wiener filter technique in terms of computational time and memory requirements. True Wiener filter requires an inversion of the full pixel covariance matrix, $\boldsymbol{C}$, which for realistic surveys would be highly non trivial. Numerical methods approximate $\boldsymbol{C}^{-1}\boldsymbol{d}$, which allows for map-level reconstruction but by itself doesn't allow calculate of the Hessian matrix for band power reconstruction. This reconstruction also requires evaluation of the determinant of the Hessian, or its derivative, trace, and where the techniques provided in this work allows orders of magnitude improvement

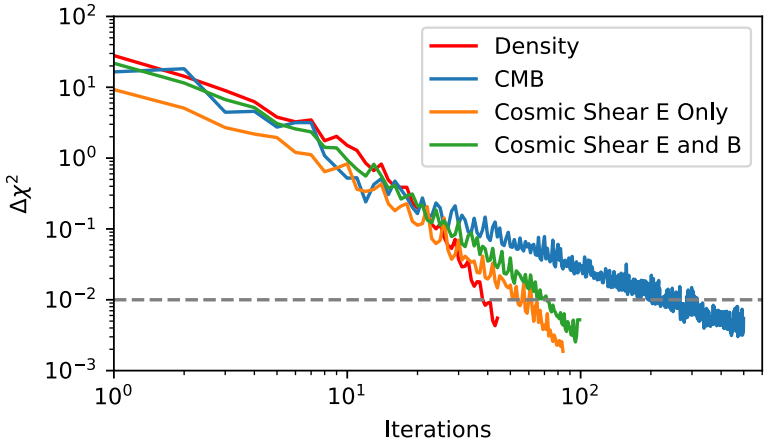


Figure 4.11: Comparison of the convergence properties of the various cosmological density fields studied in this work. All cases have 512^2 pixels, and comparable effective volume.

in realistic cases. We evaluate this determinant derivative using forward model realizations and additional optimization. This allows us to use off the shelf optimization codes such as L-BFGS, as well as convergence criteria to find the proper trade-off point between accuracy and computation time. We compare the two in Appendix 4.3.2, finding good agreement for low dimensionality problems where brute force approach is feasible.

In Figure 4.11 we show the convergence properties for the 3 cases studied in Section 4.3, as well as the joint E/B cosmic shear case presented in Appendix 4.6. While all the cases have a comparable number of unmasked pixels, convergence properties differ due to the difference of the underlying fields power-spectrum. More power at larger scales (i.e. a redder spectra) requires additional iterations to reconstruct the power within the masked regions.

While we used L-BFGS due to its well established optimization properties in very high dimensional convex optimization problems, we do not make a claim of optimality in terms of the particular technique for performing the optimization. L-BFGS constructs low rank Hessian approximation to the Hessian, which makes it a quasi-Newton method: the closer this Hessian is to the true Hessian the closer we are to true second order optimization. In this limit this method will outperform any other method, including preconditioned conjugate gradient, which is only effective if preconditioning reduces the condition number of the problem. On the other hand, true second order Newton works for any condition number. We also note that since linear problems are convex, an optimizer is always guaranteed to find the global minimum. In general, we did not find particularly large performance changes when using other optimization techniques, such as conjugate gradient. Sampling based methods, like Hamiltonian Monte Carlo, are unnecessary for these linear cases, as there is no need to

sample the distribution which is well approximated as a multivariate Gaussian, except for modes on order the size of the survey volume which are poorly sampled (and therefore do not follow the central limit theorem), in which case one can use approximations of inverse Wishart distribution developed in appendix A of [174].

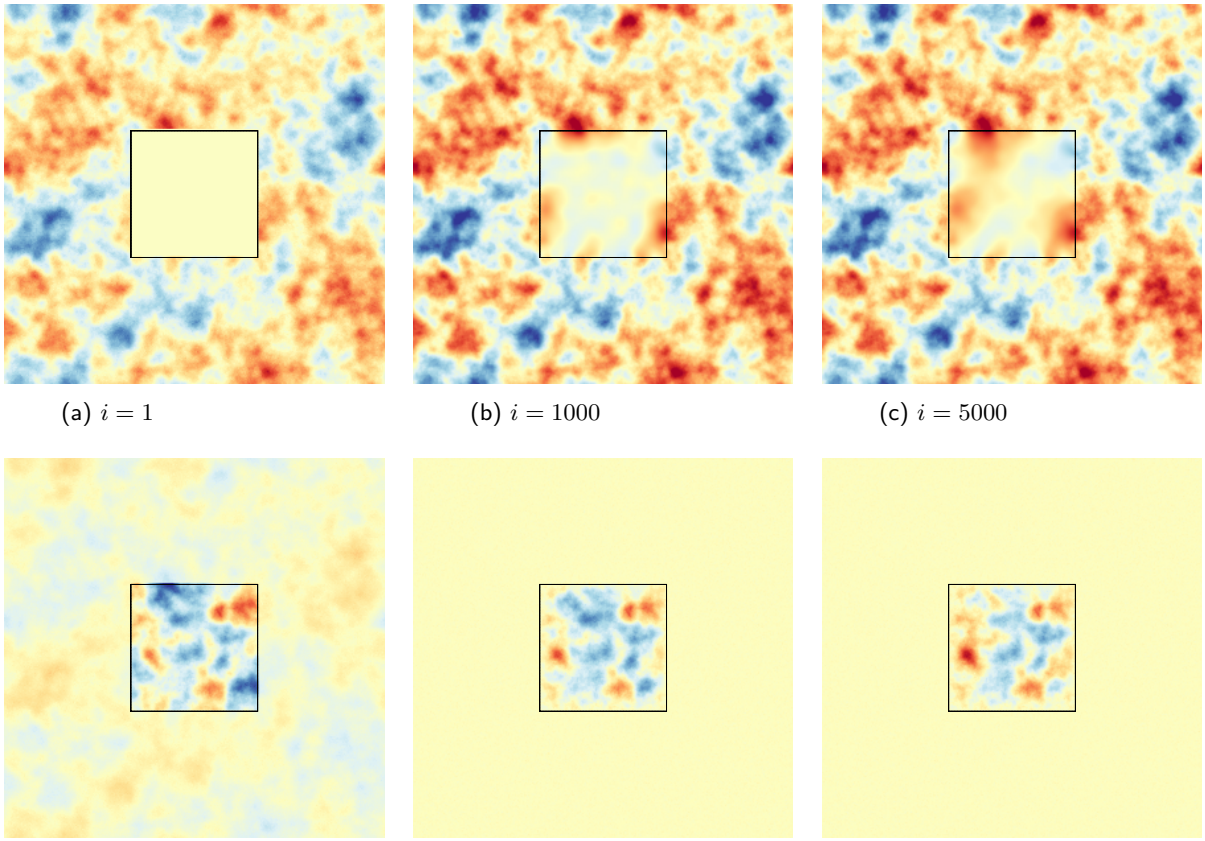
A comparison can be made in our primary CMB example in Section 4.3.3 to the results of the messenger and dual messenger field found in [108] (see Figure 6 in [108]). The L-BFGS approach requires significantly fewer iterations than the messenger field ($\sim$ 5x) and dual messenger field ($\sim$ 2x) for a comparable convergence criteria. It is possible that an optimized cooling scheme for the messenger/dual messenger field would yield similar convergence properties, but this choice would likely be problem-specific and introduce an additional parameter to tune in the optimization. In general we see little benefit in using these specialized methods over the off the shelf standard optimization methods.

Going forward, it will be useful to extend this technique to other cosmological observables such as cosmic shear tomography (such as in [180]), Lyman - α tomography, and CMB lensing. Already work has been done applying this maximum likelihood approach to CMB lensing [75, 25], and further extending this work with these methods to small scales where standard quadratic estimators [83] are known to be suboptimal [79] would be promising future approach. Another avenue of particular interest is the ability for this technique to be useful for combining multiple (biased) tracers of some underlying field to create a maximum likelihood estimate of the field. One particularly promising example is jointly maximizing the underlying density likelihood function with respect to both the shear map as well as the projected galaxy density map [192, 179].

These linear methods have a limitation when applied to nonlinear fields. Recent work [95] has demonstrated that Wiener Filtering is not optimal in terms of detecting peaks in the density field and that sparsity-based reconstruction methods can yield higher signal to noise. More general reconstruction of density fields was explored in [157], where they created MAP estimates of physical clusters with associated error estimates without making assumptions about the Gaussian of the likelihood surface. However, these approaches are ad-hoc, as the loss function they minimize cannot be theoretically justified. For the case of nonlinear large scale structure, best analog of WF reconstructions in terms of minimizing the error are the nonlinear density reconstruction techniques developed e.g. in [174, 206, 93], which first give a minimal variance reconstruction of initial gaussian density fluctuations, and then project these into the nonlinear structures using an N-body simulation.

Acknowledgments

We would like to thank Vanessa Böhm, Simone Ferraro, Stephanie Ger, Chirag Modi, and Michael Schneider for useful discussions and/or comments during the preparation of this manuscript. BH is supported by the NSF Graduate Research Fellowship, award number DGE 1106400. US acknowledges support from NSF 1814370, NSF 1839217, NASA 17-ATP17-0007 and NNX15AL17G.



(a) $i = 1$ (b) $i = 1000$ (c) $i = 5000$

Figure 4.12: *Top*: Reconstructed density field at given iteration. *Bottom*: Difference of true density field with reconstruction at each iteration. Note that we have used the same mask/noise properties, as well as color scale, as in Sec 4.3.3.

This research has made use of NASA's Astrophysics Data System. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

4.5 Convergence Criteria of CMB Reconstruction

An important question to answer is to what the required convergence criteria are for a given algorithm/observable. In general, this will depend on what sort of scales are being probed and what other sources of error exist in the problem. In this section we will consider how changing the convergence criteria, $\epsilon \equiv \Delta\chi^2$, affects the net reconstructed map. We will specialize our analysis to that of the CMB case since it has the largest condition number,

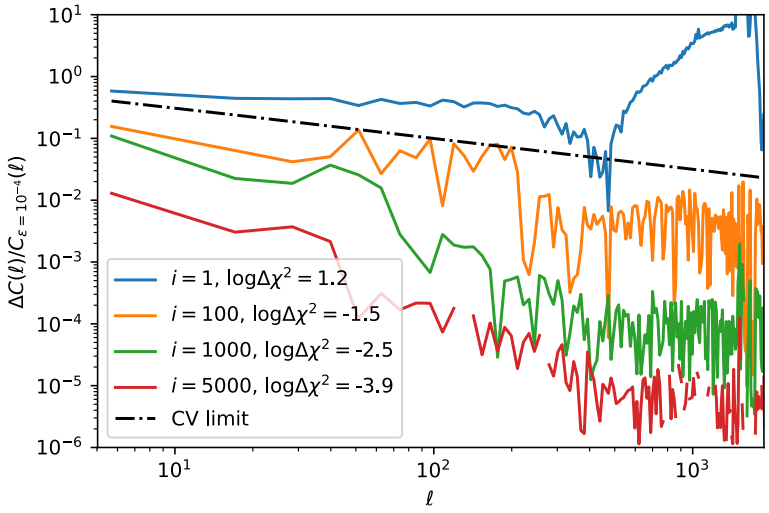
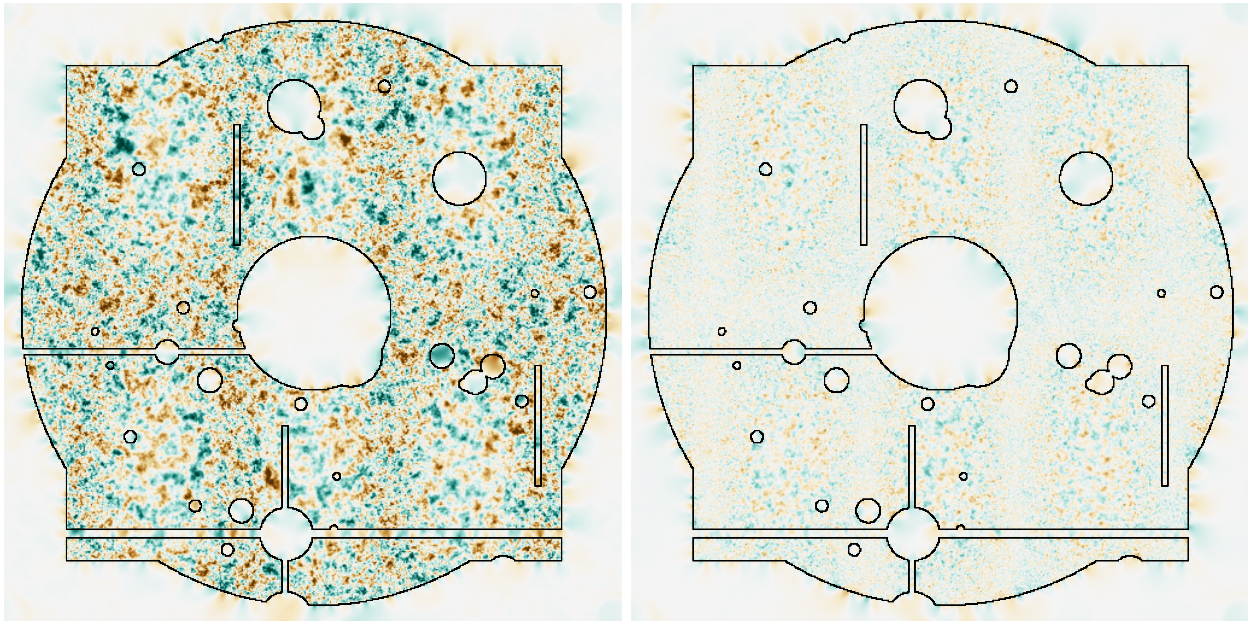


Figure 4.13: Change in convergence properties as a function of scale. We compare against the high convergence solution ($\epsilon = 10^{-4}$) rather than the true solution as the presence of noise will bias the end power spectra and calculating the noise bias and the Fisher information matrix (Hessian) for each step of the iteration would be computationally expensive.

and most power on large scales, which will be particularly sensitive to reconstruction within the masked region. A similar analysis with our cosmic shear example will lead to smaller effects.

To demonstrate the convergence properties of our technique we performed a high-accuracy run demanding $\epsilon < 10^{-4}$ as our convergence criteria, as opposed to $\epsilon < 10^{-1}$ for the runs in the main body of the paper. We show these results, as well as the difference with the true field, in Figure 4.12. Note that very quickly we find the true solution in the unmasked region, but continue to reconstruct the large scale modes in the masked region as the optimization rerouting continues.

We compare our convergence accuracy as a function of scale to the cosmic variance limit in Figure 4.13. While in practice one wants the error on the reconstruction to be well below this limit, it provides a useful guideline for the necessary accuracy for reconstruction. Note that the properties of this reconstruction are a function primarily of the survey geometry; a hypothetical full sky survey with no masked region and similar noise properties would converge much quicker to the optimal solution. Similarly, a case with smaller masked regions (for example only stellar masking) would find much faster convergence of the large scale modes.



(a) Reconstructed E Field (b) Reconstructed B Field

Figure 4.14: Maximum likelihood E and B potential fields for cosmic shear case. Noise properties, mask, and color scale are the same as in Section 4.3.4.

4.6 Joint E & B Cosmic Shear Reconstruction

In the main text we only explored reconstructing the primary (i.e. curl-free) E-mode lensing potential of cosmic shear. However, there are various potential sources of B-mode effects within realistic observed lensing maps, such as instrumental effects, clustering of source galaxies [168], and intrinsic alignments of galaxy shapes [74, 200]. To control for these effects it is useful to perform a joint optimization of both E and B modes from the shear maps. The same tools could also be applied directly to the CMB polarization field from the Q and U maps [75].

The observed shear fields γ_1 and γ_2 can be expressed in terms of the E and B potentials as

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} (\partial_x^2 - \partial_y^2) & -2\partial_x\partial_y \\ 2\partial_x\partial_y & (\partial_x^2 - \partial_y^2) \end{bmatrix} \begin{bmatrix} \phi_E \\ \phi_B \end{bmatrix}, \tag{4.26}$$

where we assume flat sky. Our response matrix now takes two signal fields (ϕ_E and ϕ_B) to two data fields (γ_1 and γ_2) and we perform the optimization over the signal fields.

To study the joint reconstruction, we use the same starting E field as in Subsection 4.3.4, but also induce a B field which has a power spectrum with the same shape as the E field but

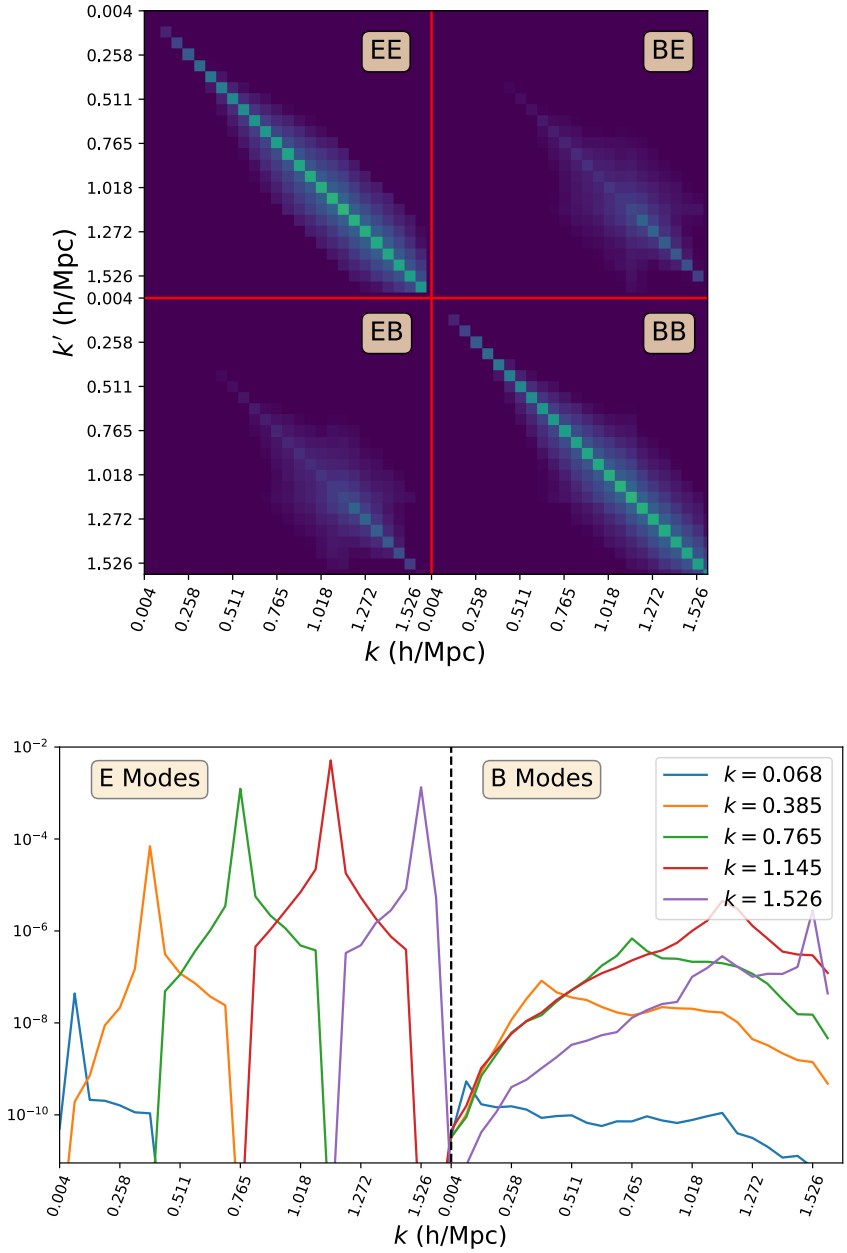


Figure 4.15: *Top*: Full two dimensional Fisher matrix for the cosmic shear E/B joint reconstruction case. The matrix can be viewed as 4 blocks, with EE and BB the response of each type of mode to itself and the BE and EB reflecting the leakage between the modes induced by the survey geometry. *Bottom*: Vertical cuts of the Fisher matrix. Note that to reduce numerical noise in the final reconstruction we have zeroed out terms in the EE/BB blocks far from the diagonal.

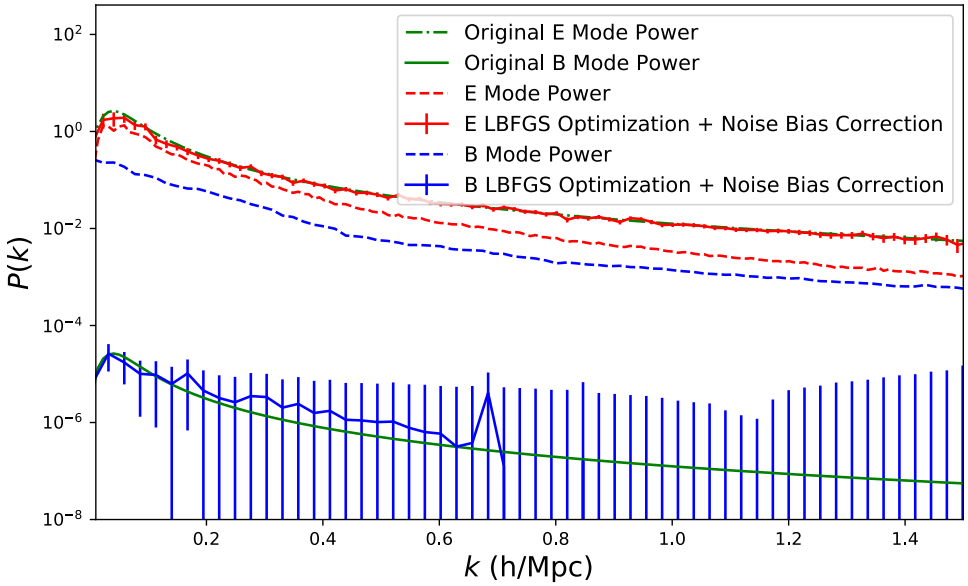


Figure 4.16: Reconstruction of the E and B power spectrum from mock observations using the maximum likelihood technique described in this work. Green lines indicate the original E/B power of the signal maps. The dashed red and blue lines indicate the power from the LBFGS optimized maps of E and B power respectively, while the solid lines indicate their MAP power spectra.

an amplitude 10^{-5} times smaller. We then do a joint reconstruction of both fields, yielding results shown in Fig 4.14.

The power spectrum estimation is now slightly more complicated as the Hessian matrix now has EE, BB, as well as EB, BE components to sum over, as shown in Fig 4.15. The EB/BE components represent leakage between the channels, which in this case is dominated by E power leaking into B. We show this power spectrum reconstruction in Figure 4.16. This is visually apparent in the reconstruction as we find an over-abundance of B power in the reconstructed map, which then gets down-weighted when this leakage is accounted for. In addition, as the B mode power is dominated by noise, it is difficult to accurately reconstruct its power from this one mock observation.

Alternatively, rather than perform a full Fisher-like analysis for the B-mode power one could instead perform multiple realizations of the B-mode leakage and average them together to form a "leakage bias" in analogous way as for the noise terms. This would have the possible advantage of requiring an additional optimization for each realization of the B-mode power, rather than an optimization for each mode of injected power as for the full Fisher analysis. If one is interested in studying many modes at once, treating the B-modes like noise bias

would be computationally expedient.

To this end, we add additional signal power to the fiducial realization, $\boldsymbol{s}_s$, as

$$\boldsymbol{s}_{l',s} = \boldsymbol{s}_s + \Delta \boldsymbol{s}_{l'}., \tag{4.27}$$

and create a noisy realization of this data using the same noise model as the observations. We choose a $\Delta \boldsymbol{s}_{l'}$ to be sufficiently small that the curvature approximation of the Fisher matrix is valid but large enough to not be susceptible to numerical artifacts. We then perform the same optimization routine to find an estimated signal vector, $\hat{\boldsymbol{s}}_{l',s+n}$, and calculate the different between this reconstructed vector and the reconstructed vector of the original field,

$$\Delta \hat{\boldsymbol{s}}_{l',s+n} = \hat{\boldsymbol{s}}_{l',s+n} - \hat{\boldsymbol{s}}_{s+n}. \tag{4.28}$$

Chapter 5

Nonlinear Reconstruction: An Approach to Modeling the $z \sim 2.5$ Cosmic Web Probed by Lyman-α Forest Tomography

> A designer knows he has achieved perfection not when there is nothing left to add, but when there is nothing left to take away.
>
> Antoine de Saint-Exupry

5.1 Introduction

A major goal of modern astrophysics is to understand how galaxies form and evolve from initial density fluctuations to the current day. Over the past few decades it has become increasingly clear that the surrounding large scale structures around galaxies play a critical role in their formation, morphology, and evolution [48, 102]. There has also been new theoretical understanding on how these large scale dark matter structures evolve, from both an analytical approach and from numerical simulations [](),a; [see 38, for an overview]. [;](),a, However, our understanding of the small scale processes driving galaxy evolution remains poor, with many competing models [38, 140]. Part of the challenge lies in the fact that most observations linking galaxy evolution and large-scale structure are at low redshifts, whereas most of the galaxy- and star-formation in the Universe peaked at the so-called 'Cosmic Noon' epoch at $z \sim 1.5-3$ [126] which remain out of reach of most large-scale structure surveys.

There are many indications of the interconnected nature of cosmic structure and galactic evolution at high redshift. Numerous studies have found that low-redshift galaxies living

in cluster environments have lower star formation rates and significantly older stellar ages than those in the field [205, 181]. This indicates that these regions underwent significant star formation and quenching at high redshift ($z > 1.5$) [199]. This is further supported by simulation work showing that protoclusters produce roughly half of their stellar content at $2 < z < 4$ and are therefore an important contribution to the overall cosmic star formation rate [33]. Beyond protoclusters, there is evidence to suggest that star formation proprieties may further depend on where the galaxy is first formed in the cluster or falls in along filamentary structure [155]. Similarly, hydrodynamical simulations [49] have suggested that the spin of galaxies may depend on the filament orientation, with simulated red and blue galaxies aligning perpendicular and parallel to the filament, respectively. Very little data is available tracing these cosmic structures at high redshift, but next generation surveys will provide the depth over sufficient sky coverage to better constrain these astrophysical processes [101, 142].

Understanding these complex relationships between baryonic properties and dark matter in the context of the overall large-scale structure environment is not only useful in modeling galaxy formation, but is also crucial in exploiting galaxies as biased tracers of large scale structure for cosmological constraints [45]. The relationships between cosmic web structures and bias has been explored in the case of tidal shear bias [12] and more recently in the case of assembly bias [159]. Quantifying the sources of bias will be needed when extending galaxy clustering surveys into the nonlinear regime where the particulars of the cosmic web may play a role [5], or in cosmic shear surveys where intrinsic alignments of galaxies will contribute substantial systematic uncertainty to precision cosmological measurements [200].

So far most studies of the cosmic web have used optically selected galaxies from spectrosopic redshift surveys as a tracer of the cosmic web. As a high number density (and threfore high spectroscopic sampling rate) is necessary for this sort of survey, this technique becomes increasingly expensive at higher redshift. The current state-of-the-art galaxy survey probing the high-redshift cosmic web is the VIPERS survey [70] on the Very Large Telescope (VLT), which has obtained redshifts for 100,000 galaxies over 24 deg^2 as the largest-ever spectroscopic campaign on that facility. This enabled a cosmic web analysis in the redshift range $0.4 < z < 1.0$ [129], which suggested segregation of massive galaxies towards filaments already at this redshift. Over the next few years, new massively-multiplexed fiber spectrographs on 8m-class telescopes, such as VLT-MOONS [34] and Subaru-PFS [PFS; 193], will allow such high-sampling rate galaxy surveys to push to $z \sim 1.5$, but would be prohibitively expensive at the "Cosmic Noon" epoch of $z \sim 2 - 3$.

In recent years, however, "intergalactic medium (IGM) tomography" [**Stark2015**, 151, 28, 117] of the hydrogen Lyα forest provides a complementary approach to mapping high-redshift large-scale structure. This technique uses dense configurations of closely-spaced star-forming galaxies, in addition to quasars, as background sources to probe the three-dimensional (3D) structure of the optically thin IGM gas at $z > 2$ on scales of several comoving Mpc. The ongoing COSMOS Lyman Alpha Mapping And Tomographic Observations (CLAMATO) survey is the first observational program to implement IGM tomography, and now has 240 sightlines covering a ~ 600 square arcmin footprint within the COSMOS

field, yielding a 3D tomographic map of the $2.05 < z < 2.55$ Lyα forest [115]. A number of $z \sim 2.3$ cosmic structures already have been detected in the CLAMATO data, including protoclusters [118] and cosmic voids [111].

In the coming years, a number of next generation spectroscopic surveys will radically increase the observational resources available for IGM tomgraphy, including the Subaru Prime Focus Spectrograph and Maunakea Spectroscopic Explorer [MSE; 135]. These telescopes will offer multiplex factors of several thousand over ~ 1 deg^2 fields of view, allowing several times the volume of the current CLAMATO data to be observed within a single night. Meanwhile, with far sparser sightline number density but significantly larger sky coverage, the Dark Energy Spectroscopic Instrument [DESI; 119] could be another interesting platform for Lyman-α forest tomography to probe large-scale over-densities. Farther into the future, the thirty-meter class facilities such as Thirty Meter Telescope [TMT; 182], Giant Magellan Telescope [GMT; 98], and European Extremely Large Telescope [EELT; 55], will have smaller fields-of-view but dramatically improved sensitivity for faint background sources at much greater sightline densities that can probe spatial scales of ~ 1 cMpc and below. The need for accurate modeling of the formation and evolution of galaxies and galaxy clusters increases in order to maximize the science return of these facilities.

The current standard procedure for IGM tomography analysis is to create a Wiener-filtered absorption map from the observed Lyα absorption features [151, 28, 117]. This absorption field can then be related to the underlying matter density through the fluctuating Gunn-Peterson approximation. This Wiener filtering does not explicitly include information about the physical processes of the system and could, in an extreme case, lead to inferred matter distributions which cannot arise from gravitational evolution. In this work, we implement a different approach, finding the maximum *a posteriori* initial density field which gives rise to the observed density field, often known as a "constrained realization." This will constrain the transmitted flux[1] field to those which are likely to arise from gravitational evolution, providing a more accurate reconstruction at $z = 2.5$. This epoch is particularly amenable to this technique since the observed structures are only mildly non-linear and have not yet undergone shell crossing. Not only will this yield information on the underlying dark matter density field, but also velocity information allowing us to deconvolve redshift space and real space quantities (see nusser:1999 for a reconstruction method applied to 1-dimensional quasar Lyα forest sightlines, and 2001Pichon for full 3D convolution). This velocity information can also help inform the astrophysical processes occurring in the region; for example combining the flux information, matter velocity information, and a galaxy catalog will provide insights into galaxy formation environmental dependence. In addition, since we have the $z = 2.5$ matter density and velocities, we are able to further evolve our field to $z = 0$ to infer the late time fate of the observed structures.

Reconstructing the initial density field has additional advantages beyond possible improvements in late time reconstruction. As there is currently no evidence for primordial

[1]It is a mild misnomer to refer to the Lyα transmission as a 'flux', but in this paper we use both terms interchangeably.

non-gaussianity [154], the power-spectrum of the initial density modes should provide a lossless statistic. The entire family of higher order correlations (such as three-point functions, density peak counts, voids, topological measures, etc.) arise due to gravitational evolution of a density field described by a single power spectra. In the case of galaxy large scale structure surveys, there has already been work towards performing this optimal reconstruction [174]. As Lyα tomography builds up toward cosmological volumes, it would be worth exploring the application of the aforementioned techniques.

In this paper we apply initial density reconstruction to mock observations of IGM Tomography using the Tomographic Absorption Reconstruction and Density inference Scheme (TARDIS). We overview the formalism in Section 5.2, describing the optimization scheme, forward model used, and measures of the cosmic web. In Section 5.3, we describe our mock data-sets which simulate Lyα tomography observations. In Section 5.4 we describe our results, and finally discuss next steps in Section 5.5.

5.2 Methodology

In order to implement our scheme to go fom observed data to the systems initial conditions we need (a) a dynamic forward model (FastPM), (b) an absorption model (FGPA), (c) mapping from field to data-space (flux skewers), (d) a noise model. In this section we describe each component of our model.

5.2.1 Modeling

Here we summarize the optimization technique and standardize notation. For a more complete description, see [171, 180, 174, 80].

We measure N skewers of flux assuming perfect identification of the continuum spectra each of length L, and stack those into a full data vector, $\boldsymbol{d}$, of total dimension $N \times L$. This data vector will depend on the initial conditions we wish to estimate at a certain resolution M, $\boldsymbol{s}$, the Lyman-α absorption model, and a noise term, $\boldsymbol{n}$, which we choose to have the same dimension as the data i.e.

$$\boldsymbol{d} = \boldsymbol{R}(\boldsymbol{s}) + \boldsymbol{n}, \tag{5.1}$$

where the $\boldsymbol{R} : M^3 \to N \times L$ is the (nonlinear) response operator composed of a forward operator and a skewer-selector function. The Gaussian information is contained in co-variance matrices, $\boldsymbol{S} = \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle$, and $\boldsymbol{N} = \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle$, for the estimated signal and noise components, which are assumed to be uncorrelated with each other, i.e. $\langle \boldsymbol{n}(\boldsymbol{R}(\boldsymbol{s}))^\dagger \rangle = 0$. In this work we are interested in maximizing the likelihood of some underlying signal given the data. The generic likelihood function can be written as

$$\begin{aligned} L(\boldsymbol{s}|\boldsymbol{d}) = (2\pi)^{-(N+M)/2} det(\boldsymbol{S}\boldsymbol{N})^{-1/2} \times \\ \exp\left[-\tfrac{1}{2}\boldsymbol{s}^\dagger \boldsymbol{S}^{-1}\boldsymbol{s} + (\boldsymbol{d} - \boldsymbol{R}(\boldsymbol{s}))^\dagger \boldsymbol{N}^{-1}(\boldsymbol{d} - \boldsymbol{R}(\boldsymbol{s}))\right], \end{aligned} \tag{5.2}$$

where we assume calculate the signal covariance $\boldsymbol{S}$ around some fiducial powerspectra. Note that the minimum variance solution for the signal field can be found by minimizing,

$$\chi^2 = \boldsymbol{s}^\dagger \boldsymbol{S}^{-1} \boldsymbol{s} + (\boldsymbol{d} - \boldsymbol{R}(\boldsymbol{s}))^\dagger \boldsymbol{N}^{-1} (\boldsymbol{d} - \boldsymbol{R}(\boldsymbol{s})), \tag{5.3}$$

with respect to $\boldsymbol{s}$. Working in quadratic order around some fixed $\boldsymbol{s_m}$ we have

$$\chi^2 = \chi_0^2 + 2\boldsymbol{g}(\boldsymbol{s} - \boldsymbol{s_m}) + (\boldsymbol{s} - \boldsymbol{s_m})\boldsymbol{D}(\boldsymbol{s} - \boldsymbol{s_m}), \tag{5.4}$$

with gradient function

$$\boldsymbol{g} = \frac{1}{2}\frac{\partial \chi^2}{\partial \boldsymbol{s}} = \boldsymbol{S}^{-1}\boldsymbol{s_m} - \boldsymbol{R'}^\dagger(\boldsymbol{s_m})\boldsymbol{N}^{-1}(\boldsymbol{d} - \boldsymbol{R}(\boldsymbol{s_m})), \tag{5.5}$$

and curvature term[2]

$$\boldsymbol{D} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial \boldsymbol{s} \partial \boldsymbol{s}} = \boldsymbol{S}^{-1} + \boldsymbol{R'}^\dagger \boldsymbol{N}^{-1} \boldsymbol{R'}. \tag{5.6}$$

Calculation of the derivative term $\boldsymbol{R'}$ requires calculation with respect to every initial mode. We use an automated differentiation framework in Appendix B of [57] o calculate Jacobian products of our evolution operator without running additional simulations. This avoids running additional involved simulations with respect to every mode, which would be prohibitively costly.

5.2.2 Optimization

As each iteration of the chain requires running a PM simulation, it is important to minimize computational time. While others have used Hamiltonian Markov Chain Monte Carlo (HMC) algorithms to find fast reconstructions for galaxy surveys [see 94, 206], in this work we are instead finding the most likely map reconstruction. We therefore use a Limited-memory Broyden Fletcher Goldfarb Shanno (LBFGS) algorithm [156], a general technique for solving nonlinear optimization problems. Rather than sampling over the entire parameter space, LBFGS takes a quasi-Newtonian approach, i.e. it is similar to the standard Newton-Ralphson method but rather than calculating the inverse of the entire Hessian (a very large matrix for a density field on the scales of interest) it iteratively updates a pseudo-Hessian as the function is being optimized.

Quasi-Newtonian methods, like L-BFGS, are only guaranteed to find extrema for convex optimization problems. For the case of large scale structure, it was demonstrated that the posterior surface is multimodal at the smallest scales but not modes probed by next generation large scale structure surveys [57]. This optimization technique was previously implemented for the case of cosmological shear measurements and CMB reconstruction, finding fast numerical conversion even in very high dimensional parameter space [80], as well as in dark-matter-only models [174, 57].

[2]Note we drop the $\boldsymbol{R''}$ term as it fluctuates with mean zero and doesn't appreciably affect the optimization.

Our implementation is based on the `vmad` framework,[3] an extension of the `abopt` framework used to perform similar reconstructions from late time galaxy fields [138]. This framework allows very fast reconstruction convergence; for cases studied in this work each reconstruction took approximately 5 CPU-hours.

5.2.3 Response Function and Forward Model

Optimization over the initial density skewers requires defining a differential forward model which will allow us to define a χ^2 problem as in Eq. 5.3 and gradient function as in Eq. 5.5.

5.2.3.1 Forward Evolution

Following the work of [57] we first use Lagrangian Perturbation Theory (LPT) to evolve the initial conditions while the field is still almost entirely linear. We do this until $z = 100.0$, at which point we then use 5 steps of FastPM [58][4] to evolve until redshift $z = 2.5$.

There are fundamental limitations due to using a particle mesh framework with limited time steps, constraints imposed by the speed requirements for optimization. As discussed in [58] and [41], halos are not fully virialized when using these methods. This will not affect our ability to reconstruct structure on $> 1\ h^{-1}\,\mathrm{Mpc}$ scales relevant for current and upcoming surveys.

We use the $z = 2.5$ particle positions to generate a density field and infer the hydrogen Lyα optical depth using the Fluctuating Gunn Peterson Approximation (FGPA), with $T = T_0(\rho/\bar{\rho})^{(\gamma-1)}$ with slope $\gamma = 1.6$ [116]. Note that we calculate the optical depth first, which is then redshift-space distorted using the inferred velocity field. Then we compute the flux $F = \exp(-\tau)$ and select lines of sight matching the positions of the mock observations. The skewers are then smoothed with a $\sigma = 1.0$ Mpc/h Gaussian filter to imitate spectrographic smoothing; this is a conservative estimate for upcoming surveys.

5.2.3.2 Overview of Forward Model

1. Initialize a Gaussian random field (the signal field).
2. Evolve field forward to $z = 2.5$ with FastPM.
3. Use FGPA to calculate a real space Lyα optical depth.
4. Use the line of sight velocity field to shift the Lyα optical depth to redshift space.
5. Exponentiate the redshift space optical depth field to get the transmitted flux field.
6. Select skewer sightlines from redshift space flux field.
7. Convolve skewers with Gaussian spectrograph smoothing.

[3] https://github.com/rainwoodman/vmad
[4] https://github.com/rainwoodman/fastpm

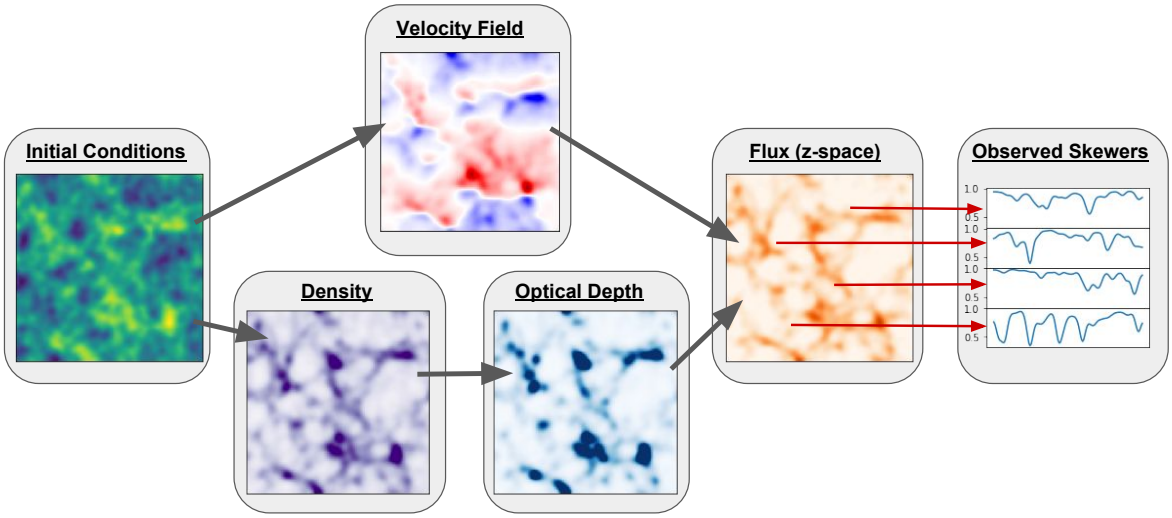


Figure 5.1: Schematic illustration of our forward model (see Sec 5.2.3.2). The underlying field we are optimizing for is the initial matter density field (left). The output of our forward model are the Lyα flux skewers probing the observational volume at the same positions as the data.

Name	N-body Code	LOS Separation (h^{-1} Mpc)	LOS Density (deg^{-2})	S/N$_{min}$ (Å^{-1})	S/N$_{max}$ (Å^{-1})	Description
T-TomoDESI	TreePM	3.7	363	1.4	4.0	Dedicated survey with DESI spectrograph (4m)
T-CLA/PFS	TreePM	2.4	863	1.4	10.0	Survey with 8-10m-class telescopes
T-30+T	TreePM	1.0	4970	2.8	10.0	Survey with 30m-class telescopes
F-CLA/PFS	FastPM	2.4	863	1.4	10.0	Same as T-CLA/PFS, but using FASTPM

Table 5.1: Mock Data Sets for Reconstructions

5.3 Mock Datasets

While the FastPM code provides a rapid convergence towards the underlying density field within the TARDIS framework, to rigorously test our reconstruction we apply the formalism to mock data generated from well-characterized large-volume, high-resolution N-body simulations. We therefore use a simulation volume run with TreePM [209, 210], which has been used for other work on Lyman-α forest tomography [**Stark2015**, 111] This simulation uses

2560^3 particles in a box with 256 $h^{-1}\,\mathrm{Mpc}$ along each dimension, with cosmological parameters $\Omega_m = 0.31$, $\Omega_b h^2 = 0.022$, $h = 0.677$, $n_s = 0.9611$, and $\sigma_8 = 0.83$. The initial conditions are generated using second order Lagrangian Perturbation Theory to $z_{ic} = 150$ and then further evolved using the TreePM code. The output was taken at $z = 2.5$ and $z = 0$ for comparison, and a $z = 2.5$ Lyman-α absorption field was generated using the FGPA with $T_0 = 2.0 \times 10^4$ and $\gamma = 1.6$.

We generated mock skewers from $(64\ h^{-1}\,\mathrm{Mpc}\,)^3$ subvolumes of the TreePM simulation with different survey parameters to mimic various ongoing and upcoming IGM tomography surveys — these are summarized in Table 5.1. The most important survey parameter is the mean sightline separation, or equivalently areal density of background sources on the sky. This is typically set by the overall sensitivity of the telescope/instrument combination and desired integration time, but in this work we simply quote the sightline separation and minimal S/N for each survey; we refer the reader to [117] for a more detailed discussion with respect to observational strategy. The CLAMATO survey [115], which is currently ongoing with the Keck-I telescope, achieves a mean separation of $2.4\ h^{-1}\,\mathrm{Mpc}$ between sightlines (albeit over a small footprint of 0.16 deg^2 at present). An IGM tomography program is currently being planned for the upcoming Prime Focus Spectrograph [188], which should achieve comparable spatial sampling as CLAMATO but over a much larger area ($\sim 15\ \mathrm{deg}^2$). Further into the 2020s, thirty-meter class telescopes such as TMT, ELT, and GMT will allow much greater sightline densities by observing fainter background sources. While the exact parameters of future IGM tomography surveys on thirty-meter telescopes will depend on instruments that are largely still under early development, for now we assume a $1\ h^{-1}\,\mathrm{Mpc}$ sightline separation. We also study a hypothetical dedicated[5] IGM tomography program carried out with the DESI spectrograph, which is currently being installed on the 4m Mayall telescope [119]. While the DESI instrument offers 5000 fibers over a 7.5 deg^2 field-of-view, we assume that 10% of the fibers will be dedicated to sky subtraction and a $1.7\times$ overhead factor in background sources targeted to maintain the specified sightline density over a finite redshift range of $\delta z = 0.3$ [117]. This implies a mean sightline separation of $3.7\ h^{-1}\,\mathrm{Mpc}$ for a dedicated DESI tomography program.

For pixel noise, we assume Gaussian random noise which varies among different skewers but is constant along each skewer. To simulate a realistic distribution of skewer S/N, we follow the prescriptions in [**Stark2015**] and [111] and draw the individual skewers' S/N from a power-law distribution with minimum value $\mathrm{S/N}_{\mathrm{min}}$ (i.e. $\boldsymbol{d}n_{\,\mathrm{los}}/\boldsymbol{d}\mathrm{S/N} \propto \mathrm{S/N}^{\alpha}$) and spectral amplitude $\alpha = 2.7$. The $\mathrm{S/N}_{\mathrm{min}}$ is the same for both the DESI and CLAMATO/PFS mocks since it reflects the actual minimal S/N in the real CLAMATO data, but for 30m-class telescopes [117] found that the S/N needs to be increased as the tomographic reconstruction is no longer limited by the shot-noise from finite skewer sampling. To be conservative, we also impose a maximal S/N for all mock datasets [115], as specified in Table 5.1.

In addition to the random pixel noise, we add continuum error to account for the diffi-

[5]There is a quasar Lyα forest component within the DESI cosmology survey, but at only $\sim 50-60\,\mathrm{deg}^{-2}$ it is far too sparse for cosmic web analysis.

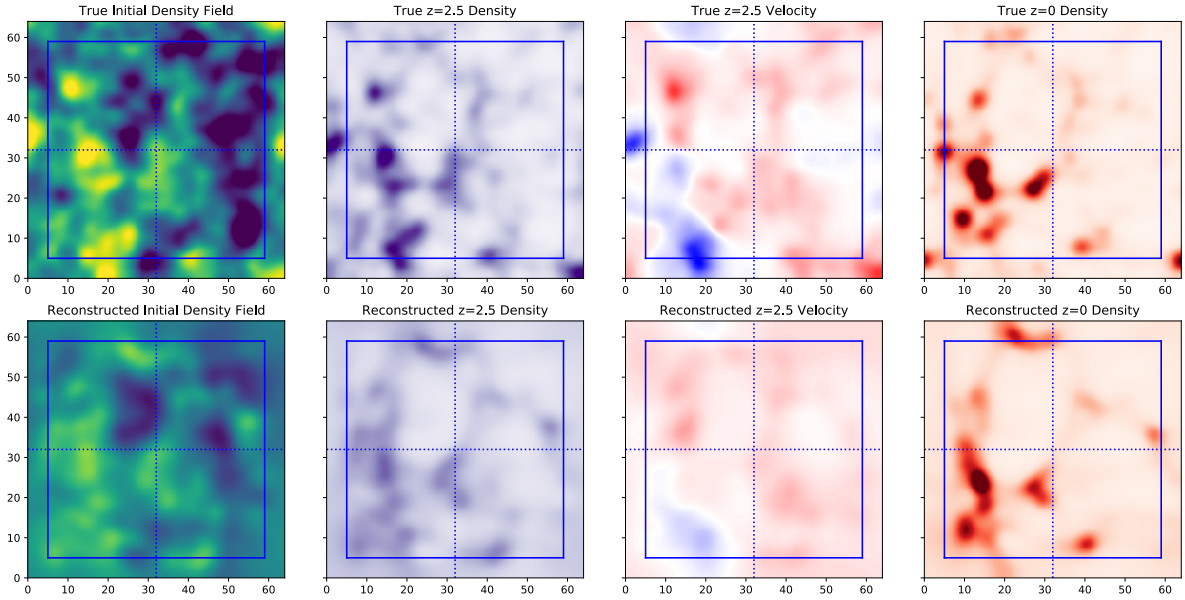


Figure 5.2: Reconstruction of various recovered quantities for the `F-CLA/PFS` mock dataset, smoothed at 2 h^{-1} Mpc , are shown on the bottom row. The true corresponding fields from the FastPM simulation are shown on the top. In all panels we project along a 5 h^{-1} Mpc slice. The region outside the solid blue box is masked in our analysis, while the dotted lines are merely to guide the eye. We find that the large scale features are qualitatively captured well in the reconstructions.

culty in identifying the intrinsic quasar or galaxy continuum. The ability to estimate the continuum is dependent on the S/N of the skewers, and we apply the fitted continuum error distribution of [111] to our mock skewers. In particular, we take our observed flux to be

$$F_{\rm obs} = \frac{F_{\rm sim}}{1+\delta_c}, \tag{5.7}$$

where δ_c is taken from an underlying Gaussian distribution with width σ_c depending on S/N along each skewer as

$$\sigma_c = \frac{0.205}{\rm S/N} + 0.015 \tag{5.8}$$

where the constants are fitted from data from the CLAMATO field. While we add continuum errors to our mock spectra, we do not directly model continuum error in TARDIS. This could be included as an off-diagonal term in the covariance matrix in future work.

In addition to the TreePM run, we have also generated mock skewers from FastPM using the exact same technique and parameters as in our forward model. This serves to isolate effects caused by known limitations of FastPM to resolve small scale halo properties, as well as provide a tool for rapid consistency checks. These are applied towards the discussions

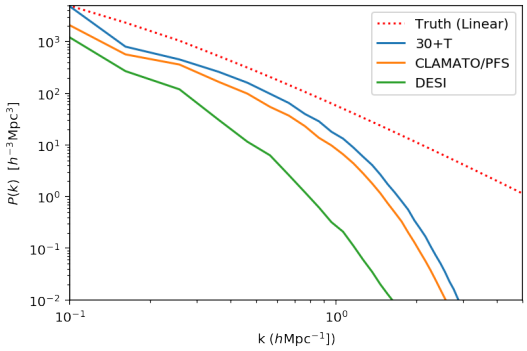


Figure 5.3: Power spectra of the reconstructed initial conditions for various experimental configurations, with the true initial conditions shown for comparison. As the number of sight-lines and spectral noise improve, power spectra reconstruction improves; however there remains a residual noise bias for realistic experiments.

regarding the code convergence in Appendix .1, and the method's sensitivity to astrophysical assumptions (Appendix .2).

5.4 Results

We apply the TARDIS method, described in §5.2, to the mock data set generated as described in §5.3. Broadly, we are interested in how well we reconstruct cosmic structures both at the observed redshift ($z = 2.5$) and the late time ($z = 0$) fate of those structures. TARDIS solves for the initial density fluctuations within the volume, which one can then use to initialize a simulation using any cosmological N-body or hydrodynamical code to study the cosmic evolution of the large-scale structure realization. For convenience, however, in this paper we continue to use FastPM to study the gravitational evolution of the TARDIS realizations at both $z = 2.5$ and $z = 0$. The $z = 2.5$ field simply the best-fit TARDIS solution, whereas to get to $z = 0$ we evolve FastPM by another five steps. We then compare the resulting fields with the 'truth' from the fiducial TreePM simulation volume.

Examples of reconstructed fields for initial density, $z = 2.5$ matter density and Lyα flux, line of sight velocity and $z = 0$ matter density for `T-CLA/PFS` are shown in Fig 5.2. In comparison with the 'true' fields, there is a strikingly good recovery of the overall filamentary backbone of the $z = 2.5$ matter density field as well as the overall distribution of the velocity field. However, the TARDIS reconstruction appears to underestimate the overall amplitude of the density field, with less prominent density peaks in both the initial conditions and $z = 2.5$ matter density. As expected, the underestimated matter power propagates through to the evolved density field at $z = 0$, where the density peaks in the reconstruction are much less prominent than the true underlying density.

The underestimated matter amplitude appears to be an artifact of the reconstruction method, and can be seen when we compare the reconstructed initial fluctuation power spectrum with that used to generate the 'true' TreePM simulation volume (Figure 5.3). There is a shortfall in the recovered power in all the mock reconstructions, especially on scales below the mean sightline density of the mock data, but also on larger scales. This gets worse with reduced sightline density of the `T-DESI` reconstruction, while conversely the improved sightline sampling of the `T-30+T` mock allows a better job of recovering the true power spectrum, although there is still a shortfall at all scales. This is possibly due to the fact that the Lyα forest absorption blends and saturates in matter overdensities. While it might be possible to correct for this reduced power in the initial density fluctuations, this is a non-trivial process which we defer to an upcoming paper that will focus on modeling galaxy protoclusters within the TARDIS framework.

Nevertheless, TARDIS appears to do a reasonable job in recovering the moderate-density cosmic web as seen in Figure 5.2. We thus focus on the large-scale cosmic web, and compare the performance of the TARDIS across cosmic time.

5.4.1 Classification of the Cosmic Web

For quantitative comparison of the large-scale structure recovery in TARDIS, we use the deformation tensor cosmic web classification of [112] and described in [114], which was inspired by [18, 72, 62]. While there exist other cosmic web classification algorithms [see summary in 29], the deformation tensor approach has a strong physical interpretation within the Zel'dovich approximation [216] and allows easy comparison to previous work in the context of Lyman-α forest tomography. However, in contrast to [114] and [112], who measured the eigenvalues and eigenvectors of Wiener-filtered maps of the Lyα transmitted flux, in this work we directly measure the eigenvalues and eigenvectors of the dark matter fields reconstructed with TARDIS, which have been first smoothed with a $R = 2\ h^{-1}\,\mathrm{Mpc}$ Gaussian kernel.

The eigenvectors and eigenvalues of the deformation tensor relate directly to the flow of matter around that point in space; matter collapses along the axis of the eigenvector when the associated eigenvalue is positive, and expands when it is negative. Points with three eigenvalues above some nonzero threshold value λ_{th} [as in 62] are nodes (roughly corresponding to (proto)clusters), two values above $\lambda_{\boldsymbol{t}_h}$ are filaments, one value above $\lambda_{\boldsymbol{t}_h}$ are sheets, and zero values above $\lambda_{\boldsymbol{t}_h}$ are voids. The deformation tensor, D_{ij}, is defined as the Hessian of the gravitational potential, Φ, i.e.

$$D_{ij} = \frac{\partial^2 \Phi}{\partial x_i \partial x_j}, \tag{5.9}$$

or equivalently in Fourier space in terms of the density field, δ_k, as

$$\tilde{D}_{ij} = \frac{k_i k_j}{k^2} \delta_k. \tag{5.10}$$

Mock Data	Pearson Coefficients			Volume Overlap (%)			
	λ_1	λ_2	λ_3	Node	Filament	Sheet	Void
T-TomoDESI	0.62	0.58	0.66	28	51	58	47
T-CLA/PFS	0.78	0.75	0.77	45	59	67	67
T-30+T	0.94	0.94	0.95	74	80	82	81

Table 5.2: Cosmic Web Recovery at $z \sim 2.5$ (Eulerian Comparison)

This tensor is then diagonalized to obtain the eigenvalues $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$ at each point on our spatial grid, ordered such that their corresponding eigenvalues are $\lambda_1 > \lambda_2 > \lambda_3$ (i.e. to demand that collapse first occurs along $\hat{e}_1$). Note that one could use the velocity field from the reconstruction itself to determine the flow at each point [e.g. 122, 144] instead of relying on the Zel'dovich approximation used in the classification here. We use the deformation tensor in order to stay consistent with past IGM tomography work [114, 112]. Cosmic web directions for our reconstructed field are thus defined by the eigenvectors with associated eigenvalues used to classify the cosmic web.

We follow [112] and [114] and define our threshhold value λ_{th} for each simulated field such that the voids occupy 21% of the total volume at $z = 2.5$ and 27% at $z = 0$ [inspired by the redshift evolution in 29]. The void fraction is somewhat arbitrary in the analysis, as long as it is consistent between the mock reconstructions and true density field used for comparison.

5.4.2 Matter/Flux Density at $z \sim 2.5$

We compare the recovery of $z = 2.5$ Lyα flux to previously-standard Wiener filtering techniques. As we are assuming the Fluctuating Gunn Peterson approximation, this reconstructed flux can be mapped directly to the density field. While past work on Wiener-filtered IGM tomographic maps [115, 28] have smoothed the field on 1.4× the mean sightline spacing, for these comparisons we smooth the respectively matter fields with a $\sigma = 2\ h^{-1}\,\mathrm{Mpc}$ Gaussian kernel. The smaller smoothing scale is appropriate for our work because our method should be better able to infer nonlinear and semi-linear structure between sight-lines. For all plots we treat the field in real space (without redshift space distortions) since our optimization is over the initial real space density field.

The reconstructed matter density fields from the various mock IGM tomography surveys (summarized in Table 5.2) are shown in the first row of Figure 5.5 in comparison with the true density field from the TreePM simulation. In all cases, they are smoothed with a $R = 2\ h^{-1}\,\mathrm{Mpc}$ Gaussian kernel. On large scales, the reconstructed density fields are well matched in terms of voids and sheets, but CLAMATO/PFS data misses out on some prominent filamentary structures and nodes as a consequence of the underestimated matter

amplitude. The 30+m telescopes, on the other hand, yield an matter density reconstruction with excellent fidelity over the entire volume.

We next calculate the characteristic eigenvalues of the deformation tensor, as described in § 5.4.1, on the smoothed matter density fields. The scatter of the eigenvalues, relative to the true underlying eigenvalues, is plotted in Figure 5.6. This reflects how well recover the amplitude of curvature of the matter density field along each cosmic web direction. The distribution of all three eigenvalues is unbiased relatively to the truth, albeit with more scatter in the case of the sparser CLAMATO/PFS reconstruction. We quantify the agreement in terms of Pearson correlation coefficients, showing the scatter from a linear trend in Table 5.1. These show a strong correlation between the reconstructed and true eigenvalues, ranging from $r = [0.78, 0.75, 0.77]$ in recovering the three eigenvalues $[\lambda_1, \lambda_2, \lambda_3]$ for CLAMATO/PFS, to the excellent reconstruction of the 30m-class telescopes with correlation coefficients of $r = [0.94, 0.94, 0.95]$.

Next, we classify the each point within the density field as void, sheet, filament, or node depending on how many of the eigenvalues are greater than the threshold value, $\lambda_i > \lambda_{\rm th}$. In the true matter density field, we find that $[22, 50, 25, 3]\%$ of the volume is occupied by voids, sheets, filaments, and nodes, respectively — by construction the reconstructed matter fields show similar volume occupation fractions to within $\pm 2\%$. The volume overlap fraction between cosmic web classifications in the mock data reconstructions compared to the true matter field are listed in Table 5.1 — these do not include a buffer region of 5 $h^{-1}\,$Mpc near the edge of the volume where we expect to be contaminated by boundary effects. For the CLAMATO/PFS mock reconstructions, the volume overlap fractions are $\sim 60 - 62\%$ for the sheets and voids, declining to 32% for the nodes. It is unsurprising that the nodes are more challenging to recover, since they occupy such a small fraction (3%) of the overall density field. These numbers are, on the surface, comparable to those found by [112] (their Table 1) for a similar CLAMATO-like mock data set, but in fact somewhat better since we are probing the matter field directly on 2 $h^{-1}\,$Mpc scales, whereas [112] were evaluating the Lyα transmission field over coarser ($4\ h^{-1}\,$Mpc) scales in the equivalent case. This improvement is due to the fact that the TARDIS incorporates the physics of gravitational evolution into its reconstructions, in contrast with Wiener-filtering, which only assumes a correlation function. The 30m-class reconstruction, as expected, fares even better thanks to its finer sightline sampling, with the voids, sheets, filaments, and nodes overlapping $[81\%, 82\%, 80\%, 74\%]$ with the true matter density cosmic web.

To further illustrate the fidelity of the recovery, Figure 5.7 shows the confusion matrix, evaluated at all the grid points in our volume, between the true cosmic web from the simulation and our reconstructions, finding good agreement. Overall, we find 80%, 60%, and 53% of the total observed volume is properly classified for `T-30+T`, `T-CLA/PFS`, and `T-DESI`, respectively. Allowing mis-classification by a structurally adjacent type (i.e. void to sheet, sheet to void/filament, filament to sheet/node, and node to filament) the agreement goes up to 98%, 96%, and 95% respectively. We also examine the eigenvector recovery by computing the dot product between the eigenvectors recovered from the reconstructions with those at

the same Cartesian point in the true matter density field[6] — with a good recovery, the recovered eigenvectors would be well aligned with the true eigenvectors and lead to dot products of order unity. These are shown in Figure 5.5. We find for $[\hat{e}_1, \hat{e}_2, \hat{e}_3]$ average alignment cosine angles of $[0.80, 0.70, 0.80]$ for `T-DESI`, $[0.87, 0.79, 0.80]$ for `T-CLA/PFS`, and $[0.96, 0.92, 0.96]$ for `T-30+T`. This is again comparable to the results derived from Wiener-filtered flux maps in [112] for the CLAMATO/PFS case, but probing smaller scales.

5.4.3 Matter Density at $z = 0$

A main motivation for the TARDIS framework is inferring the late time fate of structures and constituent galaxies found in regions observed by Lyman-α forest tomography. As output of our model, we further evolve the particle field to $z = 0$ in order to study the reconstruction. We compare this evolved field with the TreePM 'truth' at $z = 0$. The true underlying field contains cosmic structures with mass fraction $[0.15, 0.49, 0.31, 0.05]$ and volume fraction $[0.02, 0.28, 0.48, 0.22]$ for [nodes, filaments, sheets, voids], respectively.

Eulerian (real) space provides a qualitative picture of the the structures reconstructed in this limit. In Figure 5.9 (top) we show the matter field and cosmic web reconstructed for different survey mock data. While they are qualitatively similar, as described in Subsection 5.4.2, the peaks of the $z = 2.5$ density field are poorly reconstructed for realistic survey parameters. This results in significant drift of the Eulerian space structures and makes point by point comparisons difficult. This can be seen in Figure 5.9 (bottom) where the qualitative structure is quite similar, especially for 30+T, but the exact positions of nodes and filaments are in slightly different positions relative to the true matter field. This leads to unsatisfactory cosmic web recovery when evaluated in the same way as $z = 2.5$.

However, the reconstructions' cosmic web fidelity at $z = 0$ is a somewhat abstract concept since the Eulerian matter density field is not accessible via any observations. Instead, we can evaluate the reconstructed field in Lagrange space, i.e., tracking the $z = 0$ environments sampled by test particles observed at $z = 2.5$. Since we expect galaxies to act roughly like test particles in the large-scale gravitational potential, this provides a direct connection to understanding the late time fate of $z \approx 2.5$ galaxies observed in the same volume as the Lyman-α tomography data. We test this by the following: from the $z = 2.5$ density field reconstructed from the mock data reconstructions with TARDIS/FastPM, we select a set of test particles at Eulerian real-space positions $[x_{z25,i}, y_{z25,i}, z_{z25,i}]$ and track them to their $z = 0$ Eulerian positions $[x_{z0,i}, y_{z0,i}, z_{z0,i}]$ then evaluate their cosmic web eigenvalues and classifications (on the Eulerian real-space grid). From the TreePM 'true' matter density field at $z = 2.5$, we find matching test particles at the same Eulerian positions $[x_{z25,i}, y_{z25,i}, z_{z25,i}]$ and again track them to their $z = 0$ positions and environments. This process is visualized in Fig 5.8 where we show the displacement vectors for particles from the reconstructions vs. matched particles from the TreePM simulation evolved to z=0.

[6] These values only include structure in the observed region, excising an additional buffer of 2 h^{-1} Mpc near the survey boundary.

Mock Data	Pearson Coefficients			Volume Overlap (%)			
	λ_1	λ_2	λ_3	Node	Filament	Sheet	Void
`T-TomoDESI`	0.58	0.40	0.34	20	42	54	31
`T-CLA/PFS`	0.70	0.54	0.47	41	50	54	37
`T-30+T`	0.82	0.67	0.54	48	55	62	46

Table 5.3: Cosmic Web Recovery at $z = 0$ (Lagrangian Comparison)

The results from this exercise are shown in the $z = 0$ Lagrangian confusion matrix in Figure 5.10. For CLAMATO/PFS, we are able to successfully predict the $z = 0$ environment sampled by the test particles with $\sim 40 - 50\%$ fidelity, while this increases slightly to $\sim 50 - 60\%$ in the case of `T-30+T`. In both cases, $> 90\%$ of the particles are predicted to lie within ± 1 of the correct cosmic web classification, with the exception of CLAMATO/PFS node particles that are misidentified as sheet particles in 15% of cases. Nonetheless, this demonstrates the remarkable ability of TARDIS to infer the $z = 0$ environment of galaxies observed at $z = 2.5$, across over 10 Gyrs of cosmic time.

5.5 Conclusion

We present the first use of initial density reconstruction on densely-sampled Lyα forest data sets (often called "IGM tomography"), and have showed that using this technique we are able to accurately reconstruct large scale properties within the survey volume over a range of scales. In particular, we are able to recover the characterization and orientation of the cosmic web at $z = 2.5$ in terms of the deformation eigenvalues and eigenvectors assuming mock data that reflect upcoming and future multiplexed spectroscopic instrumentation. In addition, we are able to recover the qualitative structure of the observed structures at late time, $z = 0$. We have also shown that the inferred flux maps from TARDIS are more accurate and have less variance than those from Wiener filtering. Excitingly, we argue that we would be able to *predict the late-time environments of* $z \approx 2.5$ *galaxies* that are coeval with our reconstructed IGM tomography volume. This provides a promising and direct route to studying galaxies and AGN in the context of their surrounding cosmic web. For example, we would be able to identify the direct progenitors of $z = 0$ filament galaxies, and study their $z = 2.5$ galaxy properties. While we are currently limited by noise levels and sight-line spacing, in future papers we will explore ways to correct for underestimated fluctuation amplitude as a function of survey parameters.

While only explored indirectly (through $z = 0$ density reconstruction) a direct product of this technique is the particle velocity field at $z = 2.5$ which could have significant uses in informing astrophysical processes as well as cosmological constraints. For example, it

could allow accurate estimation of velocity dispersions in high-redshift protoclusters, which is currently uncertain due to challenges in disentangling galaxy peculiar motions from the large-scale Hubble expansion [207, 198, 40]. More generally, the velocity field reconstruction extends over the entire field and could be a useful addition beyond velocity fields from galaxy redshift space distortions and kinetic Sunyaev Zeldovich effects [189]. In addition, this technique would allow reconstruction of void velocity profiles which could provide constraints on modified gravity [56] and neutrino mass [133].

In this work we have held the astrophysical and cosmological parameters constant. A more complete treatment would require varying these jointly with the underlying field; however, we view this as unnecessary at this point since existing data covers a very limited volume with minimal cosmological constraining power. For next generation surveys, which will greatly expand the footprint covered, it will be required to jointly vary these parameters as well. Within the FGPA approximation the astrophysical parameters aren't a significant limitation since there are only a two global parameters of interest (A_0,γ) and our optimization scheme is fast enough that a naive Markov Chain Monte Carlo sampling would be sufficient to explore this parameter space. We explored the sensitivity of the reconstruction with respect to the absorption model in Appendix .2.

Our focus in this work is on reconstructing the moderate-density large scale structure within the survey volume, and we demonstrated that we were able to recover qualitative structure over a range of scales. Going forward, it would be useful to study how well similar techniques would be on reconstructing halo-scale (i.e. $\leq 1\ h^{-1}\,\mathrm{Mpc}$) structure, such as stacked halo and void profiles. However, going to this small scale regime reconstruction will be limited by the the specific astrophysical processes within the high-density regions where the Fluctuating Gunn Peterson Approximation will no longer hold. In particular, numerical hydrodynamic simulations have shown that there are significant deviations away from a simple temperature-density scaling relationship close to halos, in some cases even showing a turnover of the relationship [185]. It should be possible to extend the formalism proposed in this work and treating the variations from FGPA with some additional parameters to be fit for (or marginalized) in this limit, such as was done for galaxy surveys via a bias expansion [10, 107, 92]. One could also use grid-based approximation methods for baryonic effects [such as 41] to provide a more precise formation formalism for halo substructure, or use a more accurate N-body-based approximation than FGPA [186]. It would be a natural extension to test this method on mock data generated from the NyX hydrodynamic simulations designed to accurately reproduce Lyman-α absorption physics [6, 125]. Other non-Tomographic techniques have shown great promise in detecting high redshift clusters from Lyman-α observations [23], including a detection of a cluster at $2 = 2.32$ [22], but these techniques probe scales of $\approx 10\ h^{-1}\,\mathrm{Mpc}$

On the other side, additional work is needed to make this reconstruction technique useful for full scale cosmological analysis. Directly extracting power power-spectra estimates from our reconstructed maps suffers from significant noise bias effects which would make them difficult to apply directly to constrain cosmological parameters, as well as mode coupling effects due to the complexity of our forward model. Using a response formalism [as in 174,

80, 57] to estimate band-powers would be straightforward and would require $O(N)$ additional optimization runs to estimate N band-powers. However, before using these reconstructions for cosmological analysis, additional considerations are necessary, such as incorporating light-cone effects (i.e. evolution) within the survey volume and including correlated error within our model. Work in this direction is ongoing.

For future reconstruction efforts, the combination of galaxy surveys and Lyα tomographic mapping will be necessary in order to probe different redshift ranges with maximum efficiency. By including the galaxy density field in the reconstruction, we will be able to measure over-densities with higher precision than from IGM tomography alone. Furthermore, incorporating baryonic effects from hydrodynamical simulations can show how different components of the IGM trace the cosmic web at different redshifts [132]. This will allow a joint understanding of the galaxy and IGM large-scale structure distribution and how they influence each other.

Acknowledgments

We appreciate helpful discussions with Uros Seljak, Zarija Lukic, Chirag Modi, Teppei Okumura, Yu Feng, and David Spergel.

BH is supported by the NSF Graduate Research Fellowship, award number DGE 1106400, and by JSPS via the GROW program. Kavli IPMU was established by World Premier International Research Center Initiative (WPI), MEXT, Japan. This work was supported by JSPS KAKENHI Grant Number JP18H05868.

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DEC02-05CH11231.

.1 Convergence

An important question with any optimization scheme is the convergence properties of the procedure. This is particularly important for nonlinear processes like structure evolution where the likelihood surface is non-Gaussian and conceivably non-convex. We divide the issue into two questions to explore in this appendix; how many iterations are necessary for to be confident in our reconstruction technique and how sensitive is the found solution to the initial optimization starting point? For both questions we explore as a function of scale by looking at the reconstructed transfer function.

It has been shown that in the very low noise limit the likelihood surface of possible initial conditions in multi-modial; i.e. gravitational evolution is a non-injective map from initial conditions to late time structure [57]. However, this uncertainty is due to the shell-crossing degeneracy, which is only relevant for small scale non-linear structure not observed by even the optimistic configurations considered in this work. To study whether or not there is

one "true" solution or if there exist sufficiently different converged solutions, we perform the optimization analysis for the same mock catalog with different optimization starting points. In particular, we randomly choose a wide range of initial white noise fields with variance spanning three orders of magnitude. We calculated their transfer functions after 100 iterations versus a fiducial "well-converged" solution which underwent 500 iteration steps. Up to the scales of interest for the structures studied in this work, $\approx 1\ h^{-1}\,\mathrm{Mpc}$ we find very good agreement between all different starting points. There are some differences of power on very large scales, reflective of the poor constraining power of modes of order the box-size. The number density of modes per uniform bin scales as k^2, resulting in significantly more weight placed on smaller modes, until the window function (depending on the smoothing scale and sight-line density) creates a sharp cutoff. If these larger modes are of significant interest, an adiabatic optimization scheme could be used where-in the optimization begins first on a smoothed version of the observed field and then slowly small scale power is introduced back in by varying the smoothing scale as the optimization progresses [as done in 57]), or potentially directly using a multi-grid preconditioner technique [183]. Utilization of these techniques will likely be useful when extending this work for cosmological analysis.

The next important consideration is how long our scheme takes to be fully converged. We plot the transfer function as a function of convergence step in Fig .11. The exact choice of cutoff depends on the scales of interest, but since we are fundamentally limited in the transverse direction by the line of sight density and in the longitudinal direction by the spectrograph resolution, power above $k = 1.0$ h/Mpc is mostly lost to the smoothing operations on our field. By $n = 100$ we find good agreement up to $k = 1.0$ (h/Mpc) and we use this criteria as an iteration limit in the main work.

.2 Sensitivity to Cosmology and Absorption Model

In the main body of this work we have held cosmological and astrophysical parameters constant for the reconstructions. Here we briefly explore how wrong assumptions about the astrophysics or cosmology would bias our late time density field.

We use a different mock catalog, `T-IDEAL`, in order to examine the effects of varying the astrophysical parameters. This catalog has a constant signal to noise of 50 along each skewer, no continuum error, and a sightline density twice that of `T-30+T`. The idea of this super-experiment is to isolate the effects of the astrophysics from other potential sources of noise in the reconstruction. We perform our reconstructions assuming the "truth" astrophysics from our mock catalog, as well as assuming the wrong the overall flux amplitude, $A_0 = \exp(-T_0)$, and the density scaling exponent, β.

We see the effects of wrong astrophysical assumptions in Fig .12. Even with rather radically different astrophysical assumptions we find similar qualitative features in the late time structure. On the power-spectra level, we find these wrong assumptions result primarily in a bias offset from the true power-spectra. In practice, for surveys of the size studied in this work, it would be easily numerically tractable to sample over these parameters to perform the

late time reconstruction or, alternatively, to use Lyman Alpha Tomography as a constraint on these parameters.

.3 Comparison to Wiener Filtering

A promising aspect of this initial density reconstruction technique is that the reconstructed $z \sim 2$ flux field should be strictly more accurate than that from direct Wiener filtering (WF) of the skewers. This is because direct WF is a purely statistical process which does not take into account the physical evolution of the system under gravity, which further constrains the observed flux field. In this section we review the WF technique which we compare our method against. For a more general discussion of efficient WF and associated optimal bandpower construction, see [171] and [80]. For a more through description in the context of Lyman alpha forest, see [**Stark2015**].

As we are trying to reconstruct the optimal map given the data, we have to take into account both the data-data covariance, $\mathbf{C}_{\rm DD}$, the map-data covariance, $\mathbf{C}_{\rm MD}$, and the overall map noise covarance, $\mathbf{N}_{ij}$. The reconstructed map can then be expressed in terms of the observed flux, δ_F as a standard Wiener filter by ;

$$\delta_F^{\rm rec} = \mathbf{C}_{\rm MD} \cdot (\mathbf{C}_{\rm DD} + \mathbf{N})^{-1} \cdot \delta_F. \tag{11}$$

We approximate the covariance by assuming that $\mathbf{N}_{ij} = n_i^2 \delta_{ij}$ where n_i is the pixel noise. This neglects the correlated error component of continuum errors, but this is sub-dominant to the spectrograph noise and shouldn't appreciably affect our reconstructed maps. The map-data and data-data covariances are therefore approximated as

$$C = \sigma_F^2 \exp\left[-\frac{\Delta x_\perp^2}{2 l_\perp^2} - \frac{\Delta x_\parallel^2}{2 l_\parallel^2} \right]. \tag{12}$$

In order to compare directly to the Wiener filter map we use the inferred reconstructed flux map from TARDIS.

We apply the Wiener filtering algorithm to the `T-CLA/PFS` mock catalog and compare along a number of slices to the `TARDIS` reconstruction. The results are shown in Fig .13. Overall there is good agreement between all maps, with certain smaller-scale features better reconstructed in the `TARDIS` maps than the Wiener filtered maps.

A well-known feature of reconstructed maps are the presence of a bias caused by the presence of noise. We correct for this bias by a linear transformation calibrated from a separate simulated volume. The effect of this transformation is shown in Fig .14 (b). We show the reconstructed flux error in .14 (a), showing that the TARDIS maps have smaller flux error variance than the Wiener filtered maps.

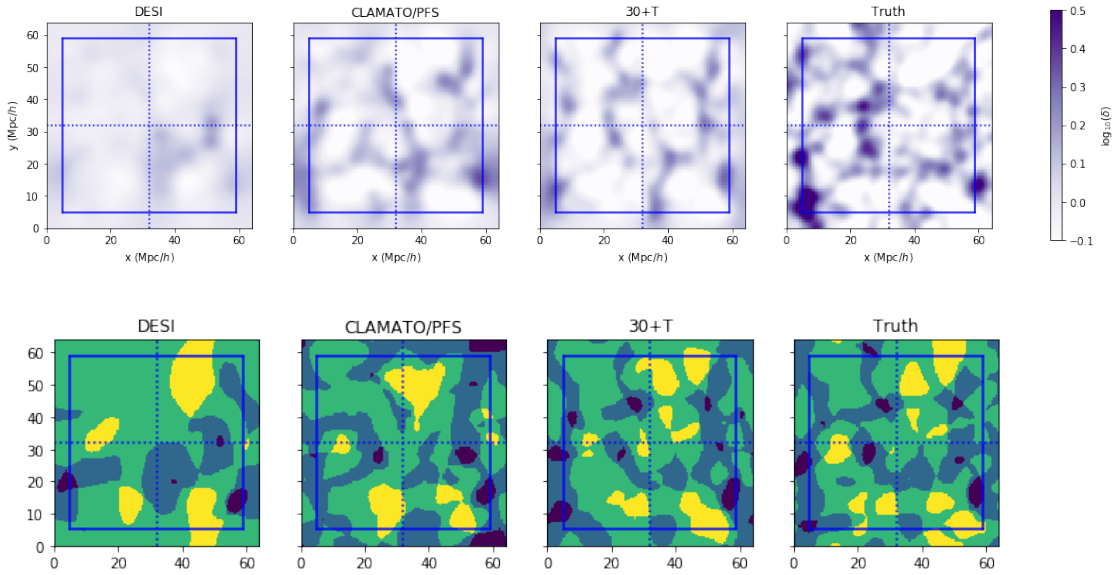


Figure 5.4: Comparison of the $z = 2.5$ reconstructed cosmic structures as classified by their eigenvalues, from `T-DESI`, `T-CLA/PFS`, and `T-30+T`, vs. the true $z = 2.5$ density field for an xy-slice. Fields have been smoothed by a $R = 2\,h^{-1}\,$Mpc Gaussian kernel. *Top*: matter density. *Bottom*: classification of cosmic structure. Dark blue indicates node, light blue indicates filament, green indicates sheet, and yellow indicates void. The region outside the solid blue box is masked in our analysis, while the dotted lines are to guide the eye.

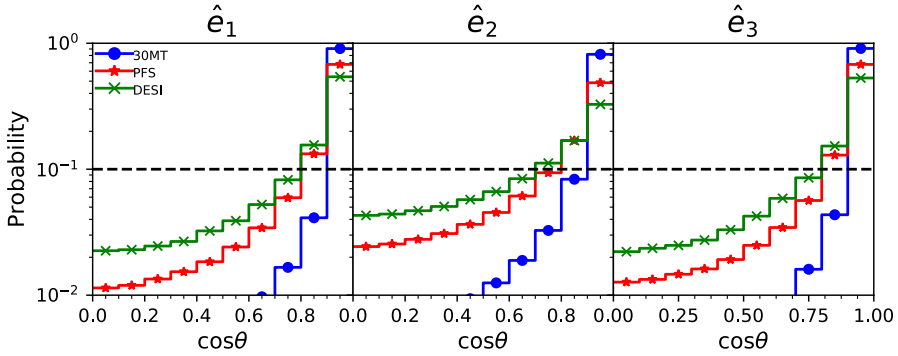


Figure 5.5: PDF showing the dot product of the eigenvectors from cosmic web reconstruction vs. the true cosmic web for various experimental configurations. $\cos\theta = 1.0$ indicates the cosmic web structures are oriented the same way, while $\cos\theta = 0.0$ indicates perpendicular alignment. Horizontal dashed line indicates the expected distribution for randomly aligned structure. In `T-30+T` the recovery of the cosmic web structure is near perfect, with only very slight misalignments on average.

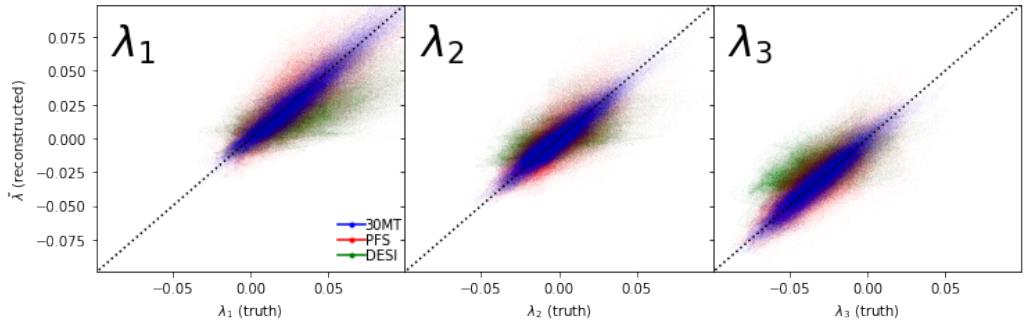


Figure 5.6: The point by point distribution of the eigenvalues inferred from the deformation tensor, smoothed by 2 h^{-1} Mpc . The magnitude of each eigenvalue indicates the magnitude of compression along the associated eigenvector. As sight-lines increase and noise decreases not only is there less scatter in the eigenvalues, but also less overall bias.

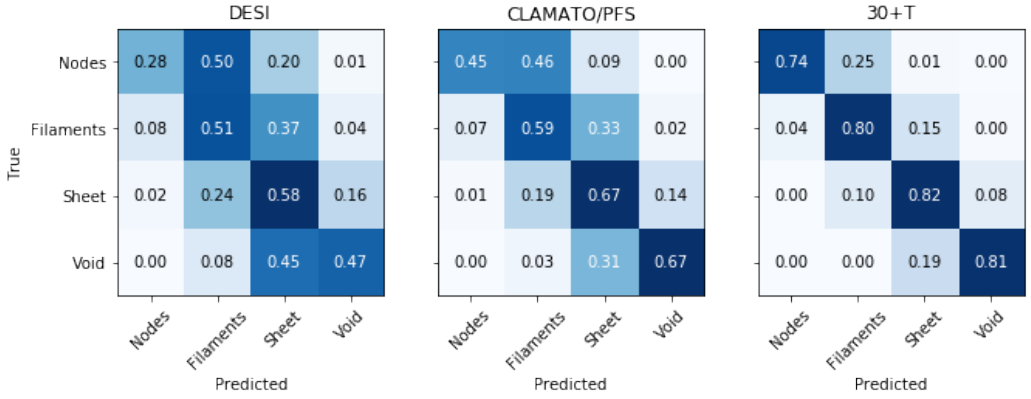


Figure 5.7: Confusion matrix for cosmic structures at $z = 2.5$ in real space showing with the reconstructed fraction printed over each cell. For `T-30+T`, we correctly identify approximately 80% of the volume.

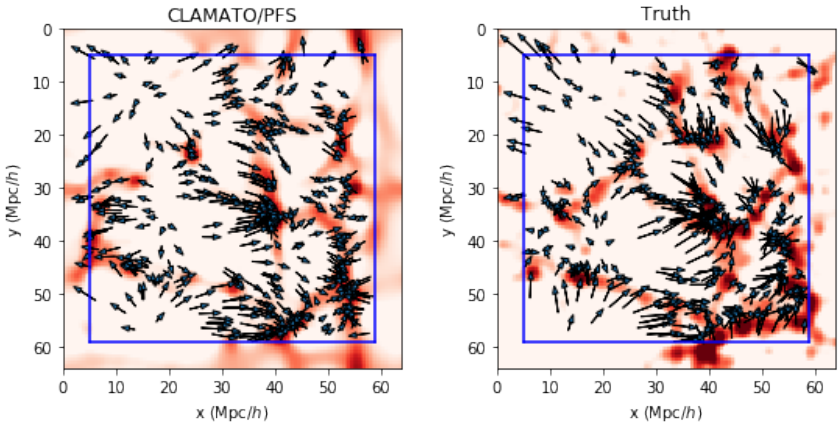


Figure 5.8: Displacement fields from $z = 2.5$ to $z = 0$ for random matched particles between the TreePM truth and the reconstructed in the mock observed volume. The underlying $z = 0$ density field is also shown. TARDIS is able to well reconstruct the movement and $z = 0$ environment of test particles identified at $z = 2.5$.

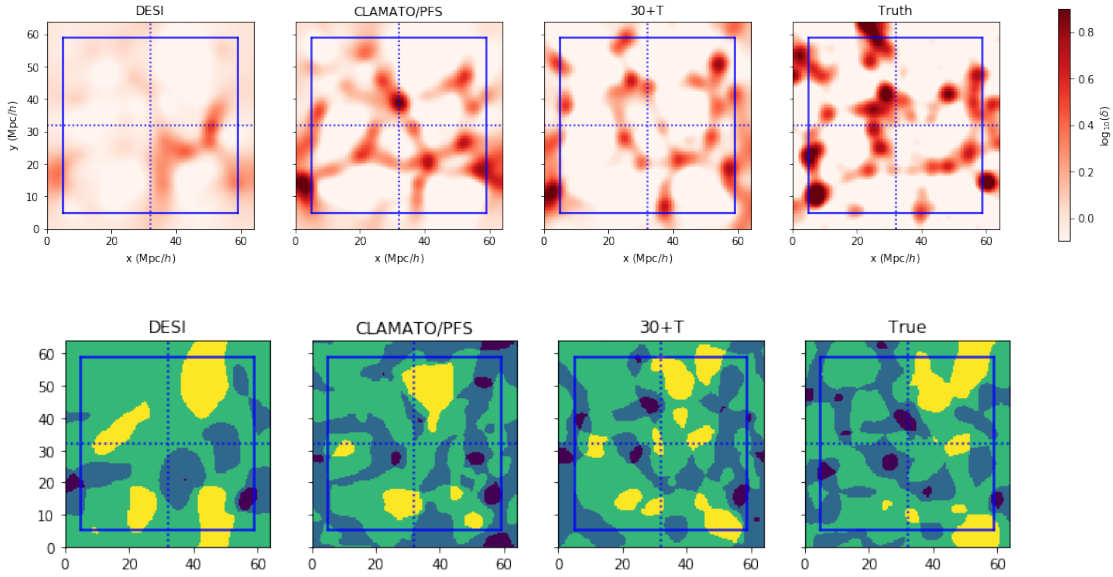


Figure 5.9: Comparison of the $z = 0$ inferred cosmic structure in Eulerian space, from `T-DESI`, `T-CLA/PFS`, and `T-30+T`, vs. the true z=0 density field. Fields have been smoothed at 2 h^{-1} Mpc . Top: matter density, Bottom: classification of cosmic structure. Dark blue indicated node, light blue indicates filament, green indicated sheet, and yellow indicates void. While the exact location of structures is poorly constrained in real space, the overall structure is quite similar especially with tight sightline spacing.

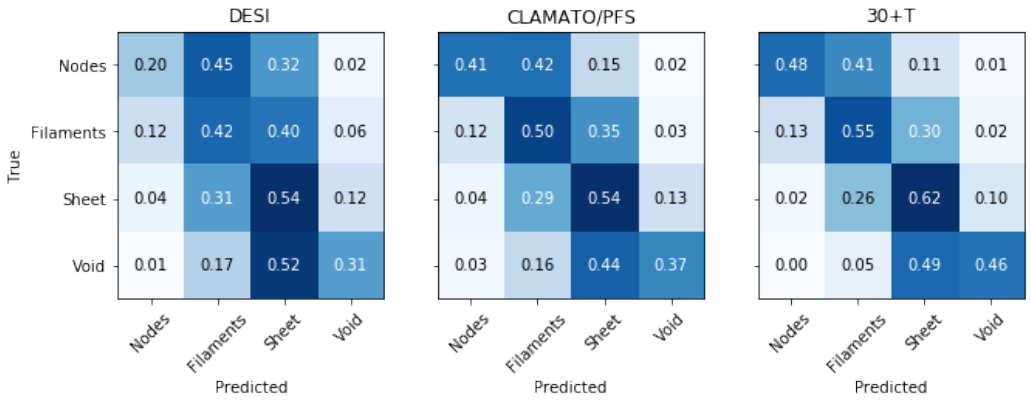


Figure 5.10: Confusion matrix for cosmic web structures at $z = 0$ in Lagrange space (i.e. comparing particles with matched in $z = 2.5$ positions) shown with the reconstructed fraction printed over each cell. While structure is not as well classified as at $z = 2.5$, classifications are approximately correct and tend toward morphologically similar environments. For comparison, the mass fraction residing in $z = 0$ nodes, filaments, sheets, and voids are $[0.15, 0.49, 0.31, 0.05]$, respectively.

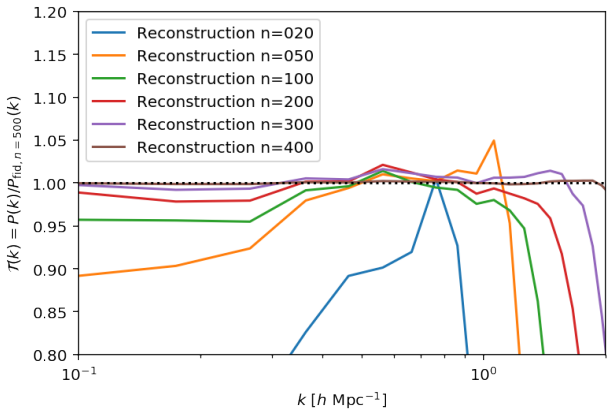


Figure .11: Transfer function with respect to a well converged solution as a function of iteration number. As the iteration number progresses, smaller and smaller scales converge. In addition, there are larger modes on order the box size that are similarly slow to converge.

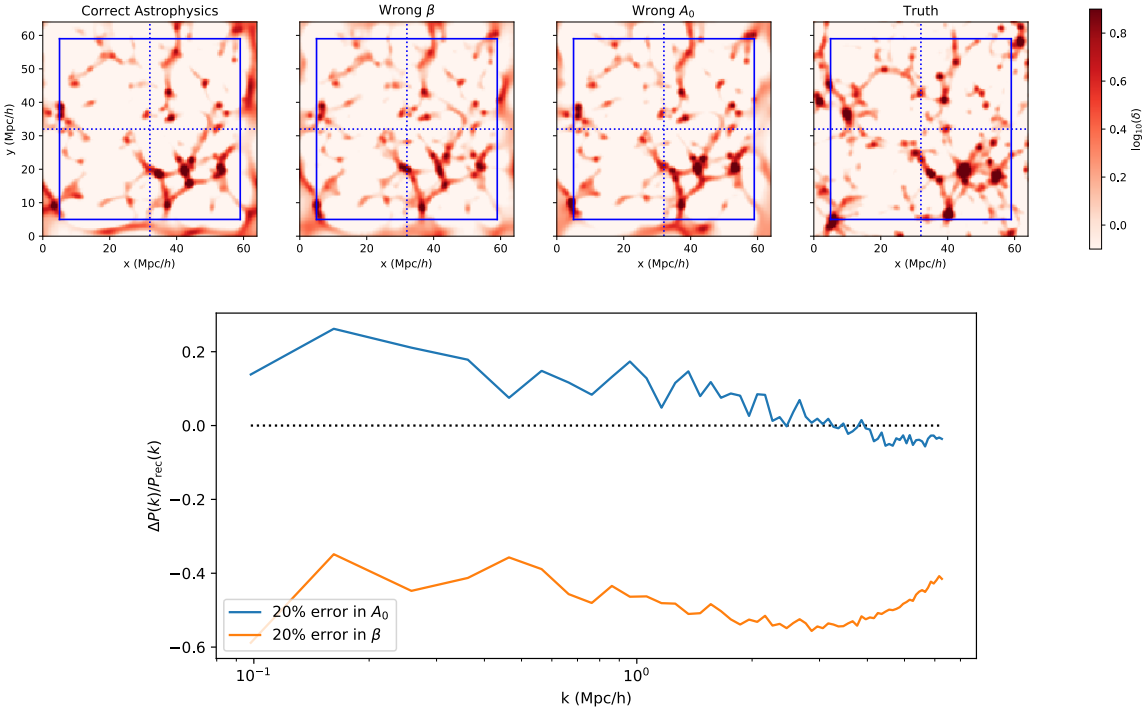


Figure .12: Effect of assuming the wrong astrophysical parameters on the $z = 0$ structure, both for a slice in real space (top) and the power spectra (bottom). Even under wrong astrophysical assumptions, we recover similar cosmic structures.

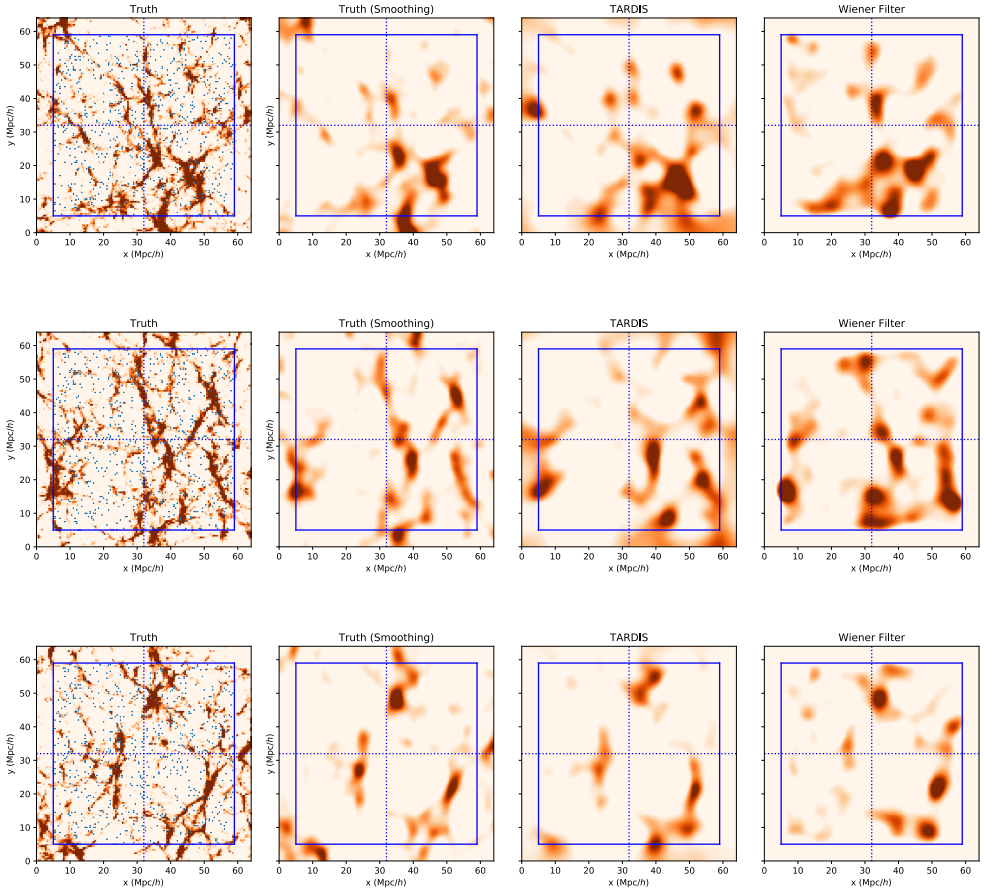
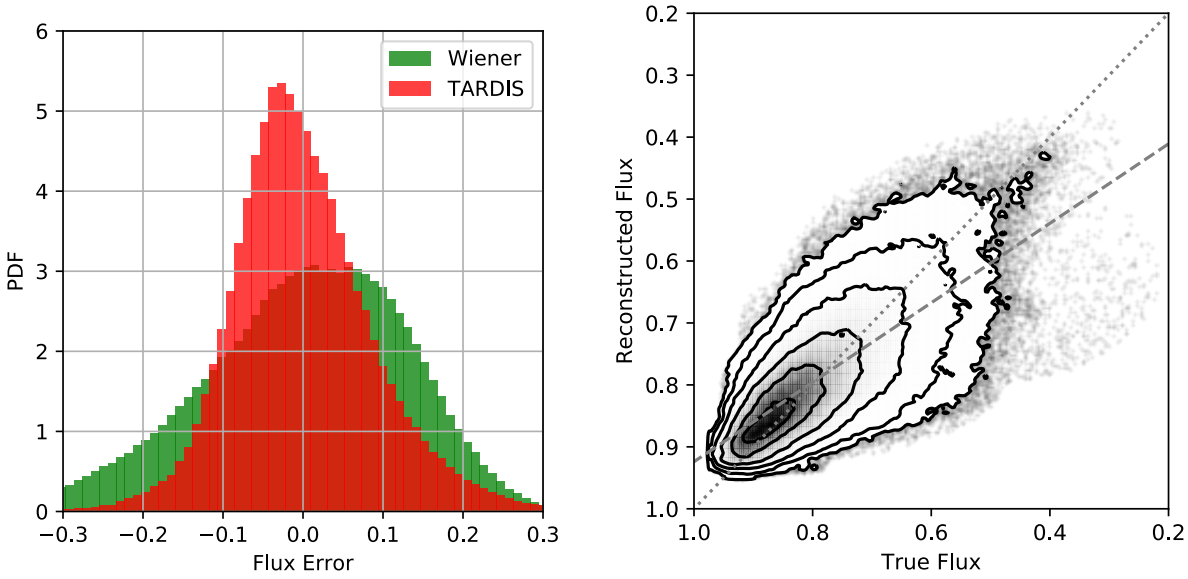


Figure .13: Comparison of the true field, TARDIS reconstructed field, and the Wiener filtered field for the T-CLA/PFS mock. In the far left panels we show the unsmoothed true flux field, with sightlines indicated as blue dots. The blue box indicates boundaries of the survey, with the blue cross to help aid the eye in matching structures. We smooth the 3 rightmost column maps on 2 h^{-1} Mpc and project over a 5 h^{-1} Mpc slice. The recovered flux field is fairly similar between TARDIS and the Wiener filter.



(a) Pixel flux error

(b) Flux reconstruction

Figure .14: Comparing the flux reconstruction for the `T-CLA/PFS` mock catalog. For these comparisons we have taken a central box which is 35 h^{-1} Mpc side-length in order to mitigate potential boundary effects and smoothed the region with a 1.5 h^{-1} Mpc Gaussian. In this plot we work in redshift space, unlike the other plots in the paper. (a) Comparison of the corrected fluxes for the Wiener filter map and `TARDIS` reconstruction vs. the true flux. (b) Scatterplot of the `TARDIS` reconstructed corrected flux vs the true flux. Also shown is the linear fit of the uncorrected flux (dashed grey line) which was linearly transformed to the $x = y$ dotted line. If interpreted as a flux PDF, each level surface indicates 0.5σ density. After this linear correction, the resulting TARDIS flux has no significant bias and mildly outperforms a linearly-corrected Wiener filtered map.

Bibliography

[1] Kevork N Abazajian et al. "CMB-S4 science book". In: *arXiv preprint arXiv:1610.02743* (2016).

[2] P. A. R. Ade et al. "Planck 2015 results. XXIV. Cosmology from Sunyaev-Zeldovich cluster counts". In: *Astron. Astrophys.* 594 (2016), A24. DOI: 10.1051/0004-6361/201525833. arXiv: 1502.01597 [astro-ph.CO].

[3] PAR Ade et al. "Planck 2015 results-XIII. Cosmological parameters". In: *Astronomy & Astrophysics* 594 (2016), A13.

[4] N Aghanim et al. "Planck 2018 results. VIII. Gravitational lensing". In: *arXiv preprint arXiv:1807.06210* (2018).

[5] S. Alam et al. "Cosmic web dependence of galaxy clustering and quenching in SDSS". In: *MNRAS* 483 (Mar. 2019), pp. 4501–4517. DOI: 10.1093/mnras/sty3477. arXiv: 1801.04878.

[6] A. S. Almgren et al. "Nyx: A Massively Parallel AMR Code for Computational Cosmology". In: *ApJ* 765, 39 (Mar. 2013), p. 39. DOI: 10.1088/0004-637X/765/1/39. arXiv: 1301.4498 [astro-ph.IM].

[7] J. Alsing, A. Heavens, and A. H. Jaffe. "Cosmological parameters, shear maps and power spectra from CFHTLenS using Bayesian hierarchical inference". In: *MNRAS* 466 (Apr. 2017), pp. 3272–3292. DOI: 10.1093/mnras/stw3161. arXiv: 1607.00008.

[8] J. Alsing et al. "Hierarchical cosmic shear power spectrum inference". In: *MNRAS* 455 (Feb. 2016), pp. 4452–4466. DOI: 10.1093/mnras/stv2501. arXiv: 1505.07840.

[9] Lauren Anderson et al. "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample". In: *Monthly Notices of the Royal Astronomical Society* 427.4 (2012), pp. 3435–3467.

[10] M. Ata, F.-S. Kitaura, and V. Müller. "Bayesian inference of cosmic density fields from non-linear, scale-dependent, and stochastic biased tracers". In: *MNRAS* 446 (Feb. 2015), pp. 4250–4259. DOI: 10.1093/mnras/stu2347. arXiv: 1408.2566.

[11] JN Bahcall and PC Joss. "Is the local supercluster a physical association". In: *The Astrophysical Journal* 203 (1976), pp. 23–32.

[12] T. Baldauf et al. "Evidence for quadratic tidal tensor bias from the halo bispectrum". In: *PRD* 86.8, 083540 (Oct. 2012), p. 083540. DOI: 10.1103/PhysRevD.86.083540. arXiv: 1201.4827 [astro-ph.CO].

[13] M. Bartelmann and P. Schneider. "Weak gravitational lensing". In: *PhysRep* 340 (Jan. 2001), pp. 291–472. DOI: 10.1016/S0370-1573(00)00082-X. eprint: astro-ph/9912508.

[14] E. J. Baxter et al. "A Measurement of CMB Cluster Lensing with SPT and DES Year 1 Data". In: (2017). arXiv: 1708.01360 [astro-ph.CO].

[15] EJ Baxter et al. "A measurement of gravitational lensing of the Cosmic Microwave Background by galaxy clusters using data from the South Pole Telescope". In: *The Astrophysical Journal* 806.2 (2015), p. 247.

[16] Francis Bernardeau et al. "Large-scale structure of the Universe and cosmological perturbation theory". In: *Physics reports* 367.1-3 (2002), pp. 1–248.

[17] V. Böhm et al. "Bayesian weak lensing tomography: Reconstructing the 3D large-scale distribution of matter with a lognormal prior". In: *PRD* 96.12, 123510 (Dec. 2017), p. 123510. DOI: 10.1103/PhysRevD.96.123510. arXiv: 1701.01886.

[18] J. R. Bond, L. Kofman, and D. Pogosyan. "How filaments of galaxies are woven into the cosmic web". In: *Nature* 380 (Apr. 1996), pp. 603–606. DOI: 10.1038/380603a0. eprint: astro-ph/9512141.

[19] FR Bouchet et al. "Perturbative Lagrangian approach to gravitational instability". In: *arXiv preprint astro-ph/9406013* (1994).

[20] UG Briel and JP Henry. "Search for X-ray filaments between galaxy clusters." In: *Astronomy and Astrophysics* 302 (1995), p. L9.

[21] E. F. Bunn, Y. Hoffman, and J. Silk. "The Wiener-filtered COBE DMR Data and Predictions for the Tenerife Experiment". In: *ApJ* 464 (June 1996), p. 1. DOI: 10.1086/177294. eprint: astro-ph/9509045.

[22] Z. Cai et al. "Mapping the Most Massive Overdensities through Hydrogen (MAMMOTH). II. Discovery of the Extremely Massive Overdensity BOSS1441 at z = 2.32". In: *ApJ* 839, 131 (Apr. 2017), p. 131. DOI: 10.3847/1538-4357/aa6a1a. arXiv: 1609.02913.

[23] Z. Cai et al. "Mapping the Most Massive Overdensity Through Hydrogen (MAMMOTH) I: Methodology". In: *ApJ* 833, 135 (Dec. 2016), p. 135. DOI: 10.3847/1538-4357/833/2/135. arXiv: 1512.06859.

[24] Ana Campos et al. "A cluster or filament of galaxies at redshift z= 2.5?" In: *The Astrophysical Journal Letters* 511.1 (1998), p. L1.

[25] J. Carron and A. Lewis. "Maximum a posteriori CMB lensing reconstruction". In: *PRD* 96.6, 063510 (Sept. 2017), p. 063510. DOI: 10.1103/PhysRevD.96.063510. arXiv: 1704.08230.

[26] Julien Carron and Antony Lewis. "Maximum a posteriori CMB lensing reconstruction". In: *arXiv preprint arXiv:1704.08230* (2017).

[27] Julien Carron, Antony Lewis, and Anthony Challinor. "Internal delensing of Planck CMB temperature and polarization". In: *Journal of Cosmology and Astroparticle Physics* 2017.05 (2017), p. 035.

[28] S. Caucci et al. "Recovering the topology of the intergalactic medium at z ˜ 2". In: *MNRAS* 386 (May 2008), pp. 211–229. DOI: 10.1111/j.1365-2966.2008.13016.x.

[29] M. Cautun et al. "Evolution of the cosmic web". In: *MNRAS* 441 (July 2014), pp. 2923–2973. DOI: 10.1093/mnras/stu768. arXiv: 1401.7866.

[30] AG Cavaliere, H Gursky, and WH Tucker. "Extragalactic X-ray sources and associations of galaxies". In: *Nature* 231.5303 (1971), pp. 437–438.

[31] Solène Chabanier, Marius Millea, and Nathalie Palanque-Delabrouille. "Matter power spectrum: from Ly α forest to CMB scales". In: *Monthly Notices of the Royal Astronomical Society* 489.2 (2019), pp. 2247–2253.

[32] Solène Chabanier et al. "The one-dimensional power spectrum from the SDSS DR14 Lyα forests". In: *Journal of Cosmology and Astroparticle Physics* 2019.07 (2019), p. 017.

[33] Y.-K. Chiang et al. "Galaxy Protoclusters as Drivers of Cosmic Star Formation History in the First 2 Gyr". In: *ApJL* 844, L23 (Aug. 2017), p. L23. DOI: 10.3847/2041-8213/aa7e7b. arXiv: 1705.01634.

[34] M. Cirasuolo et al. "MOONS: the Multi-Object Optical and Near-infrared Spectrograph for the VLT". In: *Ground-based and Airborne Instrumentation for Astronomy V*. Vol. 9147. July 2014, 91470N. DOI: 10.1117/12.2056012.

[35] Shaun Cole et al. "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications". In: *Monthly Notices of the Royal Astronomical Society* 362.2 (2005), pp. 505–534.

[36] Planck Collaboration et al. "Planck 2013 results. XVI. Cosmological parameters". In: *ASTRONOMY and ASTROPHYSICS* 571 (2013).

[37] Matthew Colless et al. "The 2df galaxy redshift survey: spectra and redshifts". In: *Monthly Notices of the Royal Astronomical Society* 328.4 (2001), pp. 1039–1063.

[38] C. J. Conselice. "The Evolution of Galaxy Structure Over Cosmic Time". In: *Ann. Rev. Astron.& Astrophys.* 52 (Aug. 2014), pp. 291–337. DOI: 10.1146/annurev-astro-081913-040037. arXiv: 1403.2783.

[39] Rupert AC Croft et al. "The Power spectrum of mass fluctuations measured from the Lyα forest at redshift z= 2.5". In: *The Astrophysical Journal* 520.1 (1999), p. 1.

[40] O. Cucciati et al. "The progeny of a Cosmic Titan: a massive multi-component proto-supercluster in formation at z=2.45 in VUDS". In: *ArXiv e-prints* (June 2018). arXiv: 1806.06073.

[41] B. Dai, Y. Feng, and U. Seljak. "A gradient based method for modeling baryons and matter in halos of fast simulations". In: *JCAP* 11, 009 (Nov. 2018), p. 009. DOI: 10.1088/1475-7516/2018/11/009. arXiv: 1804.00671.

[42] Marc Davis et al. "A survey of galaxy redshifts. II-The large scale space distribution". In: *The Astrophysical Journal* 253 (1982), pp. 423–445.

[43] Elisabetta De Filippis et al. "Measuring the three-dimensional structure of galaxy clusters. I. Application to a sample of 25 clusters". In: *The Astrophysical Journal* 625.1 (2005), p. 108.

[44] DS De Young. "Possible evidence for an intergalactic medium in clusters of galaxies". In: *The Astrophysical Journal* 173 (1972), p. L7.

[45] V. Desjacques, D. Jeong, and F. Schmidt. "Large-Scale Galaxy Bias". In: *ArXiv e-prints* (Nov. 2016). arXiv: 1611.09787.

[46] Scott Dodelson. "CMB - cluster lensing". In: *Phys. Rev.* D70 (2004), p. 023009. DOI: 10.1103/PhysRevD.70.023009. arXiv: astro-ph/0402314 [astro-ph].

[47] Scott Dodelson. *Modern cosmology*. Elsevier, 2003.

[48] A. Dressler. "Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies". In: *ApJ* 236 (Mar. 1980), pp. 351–365. DOI: 10.1086/157753.

[49] Y. Dubois et al. "Dancing in the dark: galactic properties trace spin swings along the cosmic web". In: *MNRAS* 444 (Oct. 2014), pp. 1453–1468. DOI: 10.1093/mnras/stu1227. arXiv: 1402.1165.

[50] G. Efstathiou, J. R. Bond, and S. D. M. White. "COBE background radiation anisotropies and large-scale structure in the universe". In: *MNRAS* 258 (Sept. 1992), 1P–6P. DOI: 10.1093/mnras/258.1.1P.

[51] G. Efstathiou, W. J. Sutherland, and S. J. Maddox. "The cosmological constant and cold dark matter". In: *Nature* 348 (Dec. 1990), pp. 705–707. DOI: 10.1038/348705a0.

[52] George Efstathiou and JW Eastwood. "On the clustering of particles in an expanding universe". In: *Monthly Notices of the Royal Astronomical Society* 194.3 (1981), pp. 503–525.

[53] F. Elsner and B. D. Wandelt. "Efficient Wiener filtering without preconditioning". In: *A&A* 549, A111 (Jan. 2013), A111. DOI: 10.1051/0004-6361/201220586. arXiv: 1210.4931 [astro-ph.CO].

[54] T. A. Enßlin and M. Frommert. "Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty". In: *PRD* 83.10, 105014 (May 2011), p. 105014. DOI: 10.1103/PhysRevD.83.105014. arXiv: 1002.2928 [astro-ph.IM].

[55] C. J. Evans et al. "Science case and requirements for the MOSAIC concept for a multi-object spectrograph for the European Extremely Large Telescope". In: *Ground-based and Airborne Instrumentation for Astronomy V*. Vol. 9147. July 2014, p. 914796. DOI: 10.1117/12.2055857. arXiv: 1406.6369 [astro-ph.IM].

[56] B. Falck et al. "Using voids to unscreen modified gravity". In: *MNRAS* 475 (Apr. 2018), pp. 3262–3272. DOI: 10.1093/mnras/stx3288. arXiv: 1704.08942.

[57] Y. Feng, U. Seljak, and M. Zaldarriaga. "Exploring the posterior surface of the large scale structure reconstruction". In: *JCAP* 7, 043 (July 2018), p. 043. DOI: 10.1088/1475-7516/2018/07/043. arXiv: 1804.09687.

[58] Y. Feng et al. "FASTPM: a new scheme for fast simulations of dark matter and haloes". In: *MNRAS* 463 (Dec. 2016), pp. 2273–2286. DOI: 10.1093/mnras/stw2123. arXiv: 1603.00476.

[59] BI Fesenko. "Clusters of clusters of galaxies". In: *Astrophysics* 15.4 (1979), pp. 402–405.

[60] K. B. Fisher et al. "Wiener reconstruction of density, velocity and potential fields from all-sky galaxy redshift surveys". In: *MNRAS* 272 (Feb. 1995), pp. 885–908. DOI: 10.1093/mnras/272.4.885. eprint: astro-ph/9406009.

[61] Andreu Font-Ribera, Patrick McDonald, and An$\boldsymbol{v}$ze Slosar. "How to estimate the 3D power spectrum of the Lyman-α forest". In: *Journal of Cosmology and Astroparticle Physics* 2018.01 (2018), p. 003.

[62] J. E. Forero-Romero et al. "A dynamical classification of the cosmic web". In: *MNRAS* 396 (July 2009), pp. 1815–1824. DOI: 10.1111/j.1365-2966.2009.14885.x. arXiv: 0809.4135.

[63] Riccardo Giacconi et al. "The Uhuru catalog of X-ray sources." In: *The Astrophysical Journal* 178 (1972), pp. 281–308.

[64] R Giacconi et al. "The third UHURU catalog of X-ray sources". In: *The Astrophysical Journal Supplement Series* 27 (1974), p. 37.

[65] K. M. Górski et al. "HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere". In: *ApJ* 622 (Apr. 2005), pp. 759–771. DOI: 10.1086/427976. eprint: arXiv:astro-ph/0409513.

[66] RJ Gould and GR Burbidge. "X-Rays from the Galactic Center, External Galaxies, and the Intergalactic Medium." In: *The Astrophysical Journal* 138 (1963), p. 969.

[67] Stephen A Gregory and Laird A Thompson. "The Coma/A1367 supercluster and its environs". In: (1978).

[68] James E Gunn and Bruce A Peterson. "On the Density of Neutral Hydrogen in Intergalactic Space." In: *The Astrophysical Journal* 142 (1965), pp. 1633–1641.

[69] VG Gurzadyan et al. "Probing the light speed anisotropy with respect to the cosmic microwave background radiation dipole". In: *Modern Physics Letters A* 20.01 (2005), pp. 19–28.

[70] L. Guzzo et al. "The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large-scale structure at 0.5 z 1.2". In: *A&A* 566 (June 2014), A108.

[71] Martin G Haehnelt and Max Tegmark. "Using the kinematic Sunyaev-Zeldovich effect to determine the peculiar velocities of clusters of galaxies". In: *Monthly Notices of the Royal Astronomical Society* 279.2 (1996), pp. 545–556.

[72] O. Hahn et al. "The evolution of dark matter halo properties in clusters, filaments, sheets and voids". In: *MNRAS* 381 (Oct. 2007), pp. 41–51. DOI: 10.1111/j.1365-2966.2007.12249.x. arXiv: 0704.2595.

[73] Siyu He et al. "The detection of the imprint of filaments on cosmic microwave background lensing". In: *Nature Astronomy* 2.5 (2018), p. 401.

[74] C. M. Hirata and U. Seljak. "Intrinsic alignment-lensing interference as a contaminant of cosmic shear". In: *PRD* 70.6, 063526 (Sept. 2004), p. 063526. DOI: 10.1103/PhysRevD.70.063526. eprint: astro-ph/0406275.

[75] C. M. Hirata and U. Seljak. "Reconstruction of lensing from the cosmic microwave background polarization". In: *PRD* 68.8, 083002 (Oct. 2003), p. 083002. DOI: 10.1103/PhysRevD.68.083002. eprint: astro-ph/0306354.

[76] Christopher M Hirata and U. Seljak. "Analyzing weak lensing of the cosmic microwave background using the likelihood function". In: *Physical Review D* 67.4 (2003), p. 043001.

[77] Christopher M Hirata and U. Seljak. "Reconstruction of lensing from the cosmic microwave background polarization". In: *Physical Review D* 68.8 (2003), p. 083002.

[78] Gilbert Holder and Arthur Kosowsky. "Gravitational lensing of the microwave background by galaxy clusters". In: *The Astrophysical Journal* 616.1 (2004), p. 8.

[79] B. Horowitz, S. Ferraro, and B. D. Sherwin. "Reconstructing Small Scale Lenses from the Cosmic Microwave Background". In: *ArXiv e-prints* (Oct. 2017). arXiv: 1710.10236.

[80] B. Horowitz, U. Seljak, and G. Aslanyan. "Efficient Optimal Reconstruction of Linear Fields and Band-powers from Cosmological Data". In: *ArXiv e-prints* (Sept. 2018). arXiv: 1810.00503.

[81] Benjamin Horowitz. "Revisiting primordial black holes constraints from ionization history". In: *arXiv preprint arXiv:1612.07264* (2016).

[82] Benjamin Horowitz and U. Seljak. "Cosmological constraints from thermal Sunyaev Zeldovich power spectrum revisited". In: *arXiv preprint arXiv:1609.01850* (2016).

[83] W. Hu and T. Okamoto. "Mass Reconstruction with Cosmic Microwave Background Polarization". In: *ApJ* 574 (Aug. 2002), pp. 566–574. DOI: 10.1086/341110. eprint: astro-ph/0111606.

[84] Wayne Hu, Simon DeDeo, and Chris Vale. "Cluster mass estimators from CMB temperature and polarization lensing". In: *New Journal of Physics* 9.12 (2007), p. 441.

[85] Wayne Hu and Takemi Okamoto. "Mass reconstruction with cosmic microwave background polarization". In: *The Astrophysical Journal* 574.2 (2002), p. 566.

[86] Edwin Hubble. "The distribution of extra-galactic nebulae". In: *The Astrophysical Journal* 79 (1934), p. 8.

[87] Edwin P Hubble. "Extragalactic nebulae." In: *The Astrophysical Journal* 64 (1926).

[88] Edwin P Hubble. "NGC 6822, a remote stellar system." In: *The Astrophysical Journal* 62 (1925).

[89] John P Huchra, Michael S Vogeley, and Margaret J Geller. "The CFA redshift survey: data for the south galactic CAP". In: *The Astrophysical Journal Supplement Series* 121.2 (1999), p. 287.

[90] JC Jackson. "A critique of Rees's theory of primordial gravitational radiation". In: *Monthly Notices of the Royal Astronomical Society* 156.1 (1972), 1P–5P.

[91] B. Jain and U. Seljak. "Cosmological Model Predictions for Weak Lensing: Linear and Nonlinear Regimes". In: *ApJ* 484 (July 1997), pp. 560–573. DOI: 10.1086/304372. eprint: astro-ph/9611077.

[92] J. Jasche and G. Lavaux. "Physical Bayesian modelling of the non-linear matter distribution: new insights into the Nearby Universe". In: *arXiv e-prints* (June 2018). arXiv: 1806.11117.

[93] J. Jasche and B. D. Wandelt. "Bayesian physical reconstruction of initial conditions from large-scale structure surveys". In: *MNRAS* 432 (June 2013), pp. 894–913. DOI: 10.1093/mnras/stt449. arXiv: 1203.3639.

[94] J. Jasche and B. D. Wandelt. "Bayesian physical reconstruction of initial conditions from large-scale structure surveys". In: *MNRAS* 432 (June 2013), pp. 894–913. DOI: 10.1093/mnras/stt449. arXiv: 1203.3639.

[95] N. Jeffrey et al. "Improving Weak Lensing Mass Map Reconstructions using Gaussian and Sparsity Priors: Application to DES SV". In: *ArXiv e-prints* (Jan. 2018). arXiv: 1801.08945.

[96] J. Jewell, S. Levin, and C. H. Anderson. "Application of Monte Carlo Algorithms to the Bayesian Analysis of the Cosmic Microwave Background". In: *ApJ* 609 (July 2004), pp. 1–14. DOI: 10.1086/383515. eprint: astro-ph/0209560.

[97] Mikhel Jôeveer, Jaan Einasto, and Erik Tago. "Spatial distribution of galaxies and of clusters of galaxies in the southern galactic hemisphere". In: *Monthly Notices of the Royal Astronomical Society* 185.2 (1978), pp. 357–370.

[98] M. Johns et al. "Giant Magellan Telescope: overview". In: *Ground-based and Airborne Telescopes IV*. Vol. 8444. Sept. 2012, 84441H. DOI: 10.1117/12.926716.

[99] N. Kaiser, G. Squires, and T. Broadhurst. "A Method for Weak Lensing Observations". In: *ApJ* 449 (Aug. 1995), pp. 460–+.

[100] Manoj Kaplinghat, Robert J Scherrer, and Michael S Turner. "Constraining variations in the fine-structure constant with the cosmic microwave background". In: *Physical Review D* 60.2 (1999), p. 023516.

[101] J. Kartaltepe et al. "Astro2020 Science White Paper: Assembly of the Most Massive Clusters at Cosmic Noon". In: *arXiv e-prints* (Mar. 2019). arXiv: 1903.05026.

[102] G. Kauffmann et al. "The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies". In: *MNRAS* 353 (Sept. 2004), pp. 713–731. DOI: 10.1111/j.1365-2966.2004.08117.x. eprint: astro-ph/0402030.

[103] William C Keel. *The road to galaxy formation*. Springer Science & Business Media, 2007.

[104] Woong-Tae Kim. "Heating and turbulence driving by galaxy motions in galaxy clusters". In: *The Astrophysical Journal Letters* 667.1 (2007), p. L5.

[105] RP Kirshner, A Oemler Jr, and PL Schechter. "A study of field galaxies. I-Redshifts and photometry of a complete sample of galaxies". In: *The Astronomical Journal* 83 (1978), pp. 1549–1563.

[106] F. S. Kitaura and T. A. Enßlin. "Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization". In: *MNRAS* 389 (Sept. 2008), pp. 497–544. DOI: 10.1111/j.1365-2966.2008.13341.x. arXiv: 0705.0429.

[107] F.-S. Kitaura et al. "Bayesian redshift-space distortions correction from galaxy redshift surveys". In: *MNRAS* 457 (Mar. 2016), pp. L113–L117. DOI: 10.1093/mnrasl/slw009. arXiv: 1507.08724.

[108] D. Kodi Ramanah, G. Lavaux, and B. D. Wandelt. "Wiener filter reloaded: fast signal reconstruction without preconditioning". In: *MNRAS* 468 (June 2017), pp. 1782–1793. DOI: 10.1093/mnras/stx527. arXiv: 1702.08852.

[109] E. Komatsu, D. N. Spergel, and B. D. Wandelt. "Measuring Primordial Non-Gaussianity in the Cosmic Microwave Background". In: *ApJ* 634 (Nov. 2005), pp. 14–19. DOI: 10.1086/491724. eprint: astro-ph/0305189.

[110] Eiichiro Komatsu and Uros Seljak. "The Sunyaev–Zel'dovich angular power spectrum as a probe of cosmological parameters". In: *Monthly Notices of the Royal Astronomical Society* 336.4 (2002), pp. 1256–1270.

[111] A. Krolewski et al. "Detection of z 2.3 Cosmic Voids from 3D Lyα Forest Tomography in the COSMOS Field". In: *ApJ* 861, 60 (July 2018), p. 60. DOI: 10.3847/1538-4357/aac829. arXiv: 1710.02612.

[112] A. Krolewski et al. "Measuring Alignments between Galaxies and the Cosmic Web at z
2-3 Using IGM Tomography". In: *ApJ* 837, 31 (Mar. 2017), p. 31. DOI: 10.3847/1538-4357/837/1/31. arXiv: 1612.00067.

[113] Erwin T Lau, Andrey V Kravtsov, and Daisuke Nagai. "Residual gas motions in the intracluster medium and bias in hydrostatic measurements of mass profiles of clusters". In: *The Astrophysical Journal* 705.2 (2009), p. 1129.

[114] K.-G. Lee and M. White. "Revealing the z ~ 2.5 Cosmic Web with 3D Lyα Forest Tomography: a Deformation Tensor Approach". In: *ApJ* 831, 181 (Nov. 2016), p. 181. DOI: 10.3847/0004-637X/831/2/181. arXiv: 1603.04441.

[115] K.-G. Lee et al. "First Data Release of the COSMOS Lyα Mapping and Tomography Observations: 3D Lyα Forest Tomography at 2.05z2.55". In: *ApJS* 237, 31 (Aug. 2018), p. 31. DOI: 10.3847/1538-4365/aace58. arXiv: 1710.02894.

[116] K.-G. Lee et al. "IGM Constraints from the SDSS-III/BOSS DR9 Lyα Forest Transmission Probability Distribution Function". In: *ApJ* 799, 196 (Feb. 2015), p. 196. DOI: 10.1088/0004-637X/799/2/196. arXiv: 1405.1072.

[117] K.-G. Lee et al. "Observational Requirements for Lyα Forest Tomographic Mapping of Large-scale Structure at z ~ 2". In: *ApJ* 788, 49 (June 2014), p. 49. DOI: 10.1088/0004-637X/788/1/49. arXiv: 1309.1477.

[118] K.-G. Lee et al. "Shadow of a Colossus: A z = 2.44 Galaxy Protocluster Detected in 3D Lyα Forest Tomographic Mapping of the COSMOS Field". In: *ApJ* 817, 160 (Feb. 2016), p. 160. DOI: 10.3847/0004-637X/817/2/160. arXiv: 1509.02833.

[119] M. Levi et al. "The DESI Experiment, a whitepaper for Snowmass 2013". In: *ArXiv e-prints* (Aug. 2013). arXiv: 1308.0847 [astro-ph.CO].

[120] Antony Lewis and Anthony Challinor. "Weak gravitational lensing of the CMB". In: *Physics Reports* 429.1 (2006), pp. 1–65.

[121] Antony Lewis and Lindsay King. "Cluster masses from CMB and galaxy weak lensing". In: *Physical Review D* 73.6 (2006), p. 063006.

[122] N. I. Libeskind et al. "The velocity shear tensor: tracer of halo alignment". In: *MNRAS* 428 (Jan. 2013), pp. 2489–2499. DOI: 10.1093/mnras/sts216. arXiv: 1210.4559.

[123] Anja von der Linden et al. "Weighing the Giants–I. Weak-lensing masses for 51 massive galaxy clusters: project overview, data analysis methods and cluster images". In: *Monthly Notices of the Royal Astronomical Society* 439.1 (2014), pp. 2–27.

[124] Thibaut Louis and David Alonso. "Calibrating Cluster Number Counts with CMB lensing". In: *arXiv preprint arXiv:1609.03997* (2016).

[125] Z. Lukić et al. "The Lyman α forest in optically thin hydrodynamical simulations". In: *MNRAS* 446 (Feb. 2015), pp. 3697–3724. DOI: 10.1093/mnras/stu2377. arXiv: 1406.6361.

[126] P. Madau and M. Dickinson. "Cosmic Star-Formation History". In: *Ann. Rev. Astron.& Astrophys.* 52 (Aug. 2014), pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007.

[127] Mathew Madhavacheril, Nicholas Battaglia, and Hironao Miyatake. "Fundamental Physics from Future Weak-Lensing Calibrated Sunyaev-Zel'dovich Galaxy Cluster Counts". In: (2017). arXiv: 1708.07502 [astro-ph.CO].

[128] Mathew Madhavacheril et al. "Evidence of Lensing of the Cosmic Microwave Background by Dark Matter Halos". In: *Physical review letters* 114.15 (2015), p. 151302.

[129] N. Malavasi et al. "The VIMOS Public Extragalactic Redshift Survey (VIPERS): Galaxy segregation inside filaments at $z \simeq 0.7$". In: *ArXiv e-prints* (Nov. 2016).

[130] Rachel Mandelbaum et al. "Cosmological parameter constraints from galaxy–galaxy lensing and galaxy clustering with the SDSS DR7". In: *Monthly Notices of the Royal Astronomical Society* 432.2 (2013), pp. 1544–1575.

[131] Adam B Mantz et al. "Cosmology and astrophysics from relaxed galaxy clusters–II. Cosmological constraints". In: *Monthly Notices of the Royal Astronomical Society* 440.3 (2014), pp. 2077–2098.

[132] Davide Martizzi et al. "Baryons in the CosmicWeb of IllustrisTNG - I: Gas in Knots, Filaments, Sheets and Voids". In: (2018). eprint: Arxiv:1810.01883v2. URL: http://arxiv.org/abs/1810.01883.

[133] E. Massara et al. "Voids in massive neutrino cosmologies". In: *JCAP* 11, 018 (Nov. 2015), p. 018. DOI: 10.1088/1475-7516/2015/11/018. arXiv: 1506.03088.

[134] Matteo Maturi et al. "Gravitational lensing of the CMB by galaxy clusters". In: *Astron. Astrophys.* 436 (2005), pp. 37–46. DOI: 10.1051/0004-6361:20041785. arXiv: astro-ph/0408064 [astro-ph].

[135] A. McConnachie et al. "The Detailed Science Case for the Maunakea Spectroscopic Explorer: the Composition and Dynamics of the Faint Universe". In: *arXiv e-prints* (May 2016). arXiv: 1606.00043 [astro-ph.IM].

[136] Jean-Baptiste Melin and James G Bartlett. "Measuring cluster masses with CMB lensing: a statistical approach". In: *Astronomy & Astrophysics* 578 (2015), A21.

[137] Marius Millea, Ethan Anderes, and Benjamin D. Wandelt. "Bayesian delensing of CMB temperature and polarization". In: (2017). arXiv: 1708.06753 [astro-ph.CO].

[138] C. Modi, Y. Feng, and U. Seljak. "Cosmological reconstruction from galaxy light: neural network based light-matter connection". In: *JCAP* 10, 028 (Oct. 2018), p. 028. DOI: 10.1088/1475-7516/2018/10/028. arXiv: 1805.02247.

[139] David F Mota and John D Barrow. "Local and global variations of the fine-structure constant". In: *Monthly Notices of the Royal Astronomical Society* 349.1 (2004), pp. 291–302.

[140] T. Naab and J. P. Ostriker. "Theoretical Challenges in Galaxy Formation". In: *Ann. Rev. Astron.& Astrophys.* 55 (Aug. 2017), pp. 59–109. DOI: 10.1146/annurev-astro-081913-040019. arXiv: 1612.06891.

[141] Julio F Navarro, Carlos S Frenk, and Simon DM White. "A universal density profile from hierarchical clustering". In: *The Astrophysical Journal* 490.2 (1997), p. 493.

[142] R. Overzier and N. Kashikawa. "Tracing the formation history of galaxy clusters into the epoch of reionization". In: *arXiv e-prints* (Mar. 2019). arXiv: 1903.04980.

[143] N. Padmanabhan, U. Seljak, and U. L. Pen. "Mining weak lensing surveys." In: *New Astronomy* 8 (2003), pp. 581–603. eprint: astro-ph/0210478.

[144] I. Pahwa et al. "The alignment of galaxy spin with the shear field in observations". In: *MNRAS* 457 (Mar. 2016), pp. 695–703. DOI: 10.1093/mnras/stv2930. arXiv: 1512.02236.

[145] J. Papez, L. Grigori, and R. Stompor. "Solving linear equations with messenger-field and conjugate gradients techniques - an application to CMB data analysis". In: *ArXiv e-prints* (Mar. 2018). arXiv: 1803.03462.

[146] Phillip James Edwin Peebles and Michael G Hauser. "Statistical analysis of catalogs of extragalactic objects. III. The Shane-Wirtanen and Zwicky catalogs". In: *The Astrophysical Journal Supplement Series* 28 (1974), p. 19.

[147] PJE Peebles. "Statistical analysis of catalogs of extragalactic objects. I. Theory". In: *The Astrophysical Journal* 185 (1973), pp. 413–440.

[148] Mariana Penna-Lima et al. "Calibrating the Planck Cluster Mass Scale with CLASH". In: *arXiv preprint arXiv:1608.05356* (2016).

[149] W. J. Percival et al. "The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe". In: *MNRAS* 327 (Nov. 2001), pp. 1297–1306. DOI: 10.1046/j.1365-8711.2001.04827.x. eprint: astro-ph/0105252.

[150] Saul Perlmutter et al. "Measurements of Ω and Λ from 42 high-redshift supernovae". In: *The Astrophysical Journal* 517.2 (1999), p. 565.

[151] C. Pichon et al. "Inversion of the Lyman α forest: three-dimensional investigation of the intergalactic medium". In: *MNRAS* 326 (Sept. 2001), pp. 597–620. DOI: 10.1046/j.1365-8711.2001.04595.x. eprint: astro-ph/0105196.

[152] Planck Collaboration et al. "Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters". In: *A&A* 594, A11 (Sept. 2016), A11. DOI: 10.1051/0004-6361/201526926. arXiv: 1507.02704.

[153] Planck Collaboration et al. "Planck 2015 results. XIII. Cosmological parameters". In: *A&A* 594, A13 (Sept. 2016), A13. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589.

[154] Planck Collaboration et al. "Planck 2018 results. VI. Cosmological parameters". In: *arXiv e-prints* (July 2018). arXiv: 1807.06209.

[155] S. C. Porter et al. "Star formation in galaxies falling into clusters along supercluster-scale filaments". In: *MNRAS* 388 (Aug. 2008), pp. 1152–1160. DOI: 10.1111/j.1365-2966.2008.13388.x. arXiv: 0804.4177.

[156] W. H. Press et al. *Numerical recipes in C++ : the art of scientific computing*. 2002.

[157] M. A. Price et al. " Sparse Bayesian mass-mapping with uncertainties: hypothesis testing of structure". In: (). arXiv: 1812.04014.

[158] Srinivasan Raghunathan et al. "Measuring galaxy cluster masses with CMB lensing using a Maximum Likelihood estimator: Statistical and systematic error budgets for future experiments". In: *arXiv preprint arXiv:1705.00411* (2017).

[159] S. Ramakrishnan et al. "Cosmic web anisotropy is the primary indicator of halo assembly bias". In: *arXiv e-prints* (Mar. 2019). arXiv: 1903.02007.

[160] Massimo Ramella, Margaret J Geller, and John P Huchra. "The distribution of galaxies within the'Great Wall'". In: *The Astrophysical Journal* 384 (1992), pp. 396–403.

[161] Elena Rasia et al. "Systematics in the X-ray cluster mass estimators". In: *Monthly Notices of the Royal Astronomical Society* 369.4 (2006), pp. 2013–2024.

[162] Adam G Riess et al. "Observational evidence from supernovae for an accelerating universe and a cosmological constant". In: *The Astronomical Journal* 116.3 (1998), p. 1009.

[163] Eduardo Rozo et al. "Cluster Cosmology at a Crossroads: Neutrino Masses". In: *arXiv preprint arXiv:1302.5086* (2013).

[164] G. B. Rybicki and W. H. Press. "Interpolation, realization, and reconstruction of noisy, irregularly sampled data". In: *ApJ* 398 (Oct. 1992), pp. 169–176. DOI: 10.1086/171845.

[165] Craig L Sarazin. "X-ray emission from clusters of galaxies". In: *Reviews of Modern Physics* 58.1 (1986), p. 1.

[166] Caleb Scharf et al. "Evidence for X-ray emission from a large-scale filament of galaxies?" In: *The Astrophysical Journal Letters* 528.2 (1999), p. L73.

[167] M. Schmidt. "Large Redshifts of Five Quasi-Stellar Sources." In: *ApJ* 141 (Apr. 1965), p. 1295. DOI: 10.1086/148217.

[168] P. Schneider, L. van Waerbeke, and Y. Mellier. "B-modes in cosmic shear from source redshift clustering". In: *A&A* 389 (July 2002), pp. 729–741. DOI: 10.1051/0004-6361:20020626. eprint: astro-ph/0112441.

[169] Michael Seldner et al. "New reduction of the Lick catalog of galaxies". In: *The Astronomical Journal* 82 (1977), pp. 249–256.

[170] U. Seljak. "Bias, redshift space distortions and primordial nongaussianity of nonlinear transformations: application to Ly-α forest". In: *Journal of Cosmology and Astroparticle Physics* 2012.03 (2012), p. 004.

[171] U. Seljak. "Cosmography and Power Spectrum Estimation: A Unified Approach". In: *ApJ* 503 (Aug. 1998), pp. 492–501. DOI: 10.1086/306019. eprint: astro-ph/9710269.

[172] U. Seljak and Matias Zaldarriaga. "Lensing-induced cluster signatures in the cosmic microwave background". In: *The Astrophysical Journal* 538.1 (2000), p. 57.

[173] Uros Seljak. "Gravitational lensing effect on cosmic microwave background anisotropies: A Power spectrum approach". In: *arXiv preprint astro-ph/9505109* (1995).

[174] U. Seljak et al. "Towards optimal extraction of cosmological information from nonlinear data". In: *ArXiv e-prints* (June 2017). arXiv: 1706.06645.

[175] D. S. Seljebotn et al. "Multi-resolution Bayesian CMB component separation through Wiener-filtering with a pseudo-inverse preconditioner". In: *ArXiv e-prints* (Oct. 2017). arXiv: 1710.00621 [astro-ph.IM].

[176] CD Shane and CA Wirtanen. In: *Lick Obs* 22.1 (1967).

[177] Blake D Sherwin et al. "Two-season Atacama Cosmology Telescope polarimeter lensing power spectrum". In: *Physical Review D* 95.12 (2017), p. 123529.

[178] G Simard et al. "Constraints on Cosmological Parameters from the Angular Power Spectrum of a Combined 2500 deg2 SPT-SZ and Planck Gravitational Lensing Map". In: *The Astrophysical Journal* 860.2 (2018), p. 137.

[179] P. Simon. "Improving three-dimensional mass mapping with weak gravitational lensing using galaxy clustering". In: *A&A* 560, A33 (Dec. 2013), A33. DOI: 10.1051/0004-6361/201219257. arXiv: 1203.6205.

[180] P. Simon, A. N. Taylor, and J. Hartlap. "Unfolding the matter distribution using three-dimensional weak gravitational lensing". In: *MNRAS* 399 (Oct. 2009), pp. 48–68. DOI: 10.1111/j.1365-2966.2009.15246.x. arXiv: 0907.0016.

[181] R. A. Skibba et al. "Galaxy Zoo: disentangling the environmental dependence of morphology and colour". In: *MNRAS* 399 (Oct. 2009), pp. 966–982. DOI: 10.1111/j.1365-2966.2009.15334.x. arXiv: 0811.3970.

[182] W. Skidmore, TMT International Science Development Teams, and T. Science Advisory Committee. "Thirty Meter Telescope Detailed Science Case: 2015". In: *Research in Astronomy and Astrophysics* 15, 1945 (Dec. 2015), p. 1945. DOI: 10.1088/1674-4527/15/12/001. arXiv: 1505.01195 [astro-ph.IM].

[183] K. M. Smith, O. Zahn, and O. Doré. "Detection of gravitational lensing in the cosmic microwave background". In: *PRD* 76.4, 043510 (Aug. 2007), p. 043510. DOI: 10.1103/PhysRevD.76.043510. arXiv: 0705.3980.

[184] Kendrick M Smith et al. "Delensing CMB polarization with external datasets". In: *Journal of Cosmology and Astroparticle Physics* 2012.06 (2012), p. 014.

[185] D. Sorini et al. "A Fundamental Test for Galaxy Formation Models: Matching the Lyman-α Absorption Profiles of Galactic Halos Over Three Decades in Distance". In: *ApJ* 859, 125 (June 2018), p. 125. DOI: 10.3847/1538-4357/aabb52. arXiv: 1709.03988.

[186] D. Sorini et al. "Modeling the Lyα Forest in Collisionless Simulations". In: *ApJ* 827, 97 (Aug. 2016), p. 97. DOI: 10.3847/0004-637X/827/2/97. arXiv: 1602.08099.

[187] KT Story et al. "A measurement of the cosmic microwave background gravitational lensing potential from 100 square degrees of SPTpol data". In: *The Astrophysical Journal* 810.1 (2015), p. 50.

[188] H. Sugai et al. "Prime Focus Spectrograph for the Subaru telescope: massively multiplexed optical and near-infrared fiber spectrograph". In: *Journal of Astronomical Telescopes, Instruments, and Systems* 1.3, 035001 (July 2015), p. 035001. DOI: 10.1117/1.JATIS.1.3.035001. arXiv: 1507.00725 [astro-ph.IM].

[189] N. S. Sugiyama, T. Okumura, and D. N. Spergel. "Will kinematic Sunyaev-Zel'dovich measurements enhance the science return from galaxy redshift surveys?" In: *JCAP* 1, 057 (Jan. 2017), p. 057. DOI: 10.1088/1475-7516/2017/01/057. arXiv: 1606.06367.

[190] R. A. Sunyaev and Zeldovich. "Small scale fluctuations of relic radiation". In: *Astrophys. Space Sci.* 7 (1970), pp. 3–19.

[191] RA Sunyaev and YA B Zeldovich. "Comments Astrophys". In: *Space Phys* 4.173 (1972), p. 9.

[192] R. M. Szepietowski et al. "Density mapping with weak lensing and phase information". In: *MNRAS* 440 (May 2014), pp. 2191–2200. DOI: 10.1093/mnras/stu380. arXiv: 1306.5324.

[193] M. Takada et al. "Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph". In: *ApJ* 66, R1 (Feb. 2014), R1. DOI: 10.1093/pasj/pst019. arXiv: 1206.0737.

[194] M. Tegmark. "How to measure CMB power spectra without losing information". In: *PRD* 55 (May 1997), pp. 5895–5907. eprint: astro-ph/9611174.

[195] M. Tegmark, A. de Oliveira-Costa, and A. J. Hamilton. "High resolution foreground cleaned CMB map from WMAP". In: *PRD* 68.12, 123523 (Dec. 2003), p. 123523. DOI: 10.1103/PhysRevD.68.123523. eprint: astro-ph/0302496.

[196] Max Tegmark et al. "Cosmological constraints from the SDSS luminous red galaxies". In: *Physical Review D* 74.12 (2006), p. 123507.

[197] Eric R Tittley and Mark Henriksen. "A filament between galaxy clusters A3391 and A3395". In: *The Astrophysical Journal* 563.2 (2001), p. 673.

[198] M. W. Topping et al. "Understanding Large-scale Structure in the SSA22 Protocluster Region Using Cosmological Simulations". In: *ApJ* 852, 134 (Jan. 2018), p. 134. DOI: 10.3847/1538-4357/aa9f0f. arXiv: 1709.06572.

[199] K.-V. H. Tran et al. "Reversal of Fortune: Confirmation of an Increasing Star Formation-Density Relation in a Cluster at z = 1.62". In: *ApJL* 719 (Aug. 2010), pp. L126–L129. DOI: 10.1088/2041-8205/719/2/L126. arXiv: 1005.5126.

[200] M. A. Troxel and M. Ishak. "The intrinsic alignment of galaxies and its impact on weak gravitational lensing in an era of precision cosmology". In: *PhysRep* 558 (Feb. 2015), pp. 1–59. DOI: 10.1016/j.physrep.2014.11.001. arXiv: 1407.6990.

[201] M. A. Troxel et al. "Dark Energy Survey Year 1 Results: Cosmological Constraints from Cosmic Shear". In: *ArXiv e-prints* (Aug. 2017). arXiv: 1708.01538.

[202] Dmitriy Tseliakhovich and Christopher Hirata. "Relative velocity of dark matter and baryonic fluids and the formation of the first structures". In: *Physical Review D* 82.8 (2010), p. 083520.

[203] Chris Vale, Alexandre Amblard, and Martin White. "Cluster lensing of the CMB". In: *New Astronomy* 10.1 (2004), pp. 1–15.

[204] L. Van Waerbeke et al. "Cosmic shear statistics and cosmology". In: *A&A* 374 (Aug. 2001), pp. 757–769. DOI: 10.1051/0004-6361:20010766. eprint: astro-ph/0101511.

[205] D. A. Wake et al. "The Environmental Dependence of Galaxy Colors in Intermediate-Redshift X-Ray-selected Clusters". In: *ApJ* 627 (July 2005), pp. 186–202. DOI: 10.1086/430117. eprint: astro-ph/0503480.

[206] H. Wang et al. "ELUCID-Exploring the Local Universe with the Reconstructed Initial Density Field. I. Hamiltonian Markov Chain Monte Carlo Method with Particle Mesh Dynamics". In: *ApJ* 794, 94 (Oct. 2014), p. 94. DOI: 10.1088/0004-637X/794/1/94. arXiv: 1407.3451.

[207] T. Wang et al. "Discovery of a Galaxy Cluster with a Violently Starbursting Core at z = 2.506". In: *ApJ* 828, 56 (Sept. 2016), p. 56. DOI: 10.3847/0004-637X/828/1/56. arXiv: 1604.07404.

[208] JK Webb et al. "Indications of a spatial variation of the fine structure constant". In: *Physical Review Letters* 107.19 (2011), p. 191101.

[209] M. White. "The Mass Function". In: *ApJS* 143 (Dec. 2002), pp. 241–255. DOI: 10.1086/342752. eprint: astro-ph/0207185.

[210] M. White et al. "Particle Mesh Simulations of the Lyα Forest and the Signature of Baryon Acoustic Oscillations in the Intergalactic Medium". In: *ApJ* 713 (Apr. 2010), pp. 383–393. DOI: 10.1088/0004-637X/713/1/383. arXiv: 0911.5341.

[211] Norbert Wiener. "Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications". In: (1949).

[212] Jaiyul Yoo and Matias Zaldarriaga. "Improved estimation of cluster mass profiles from the cosmic microwave background". In: *Physical Review D* 78.8 (2008), p. 083002.

[213] Jaiyul Yoo, Matias Zaldarriaga, and Lars Hernquist. "Lensing reconstruction of cluster-mass cross correlation with cosmic microwave background polarization". In: *Physical Review D* 81.12 (2010), p. 123006.

[214] Donald G York et al. "The sloan digital sky survey: Technical summary". In: *The Astronomical Journal* 120.3 (2000), p. 1579.

[215] Matias Zaldarriaga and U. Seljak. "Gravitational lensing effect on cosmic microwave background polarization". In: *Physical Review D* 58.2 (1998), p. 023003.

[216] Y. B. Zel'dovich. "Gravitational instability: An approximate theory for large density perturbations." In: *A&A* 5 (Mar. 1970), pp. 84–89.

[217] Fritz Zwicky. "Compact and Dispersed Cosmic Matter, Part I". In: *Advances in Astronomy and Astrophysics*. Vol. 5. Elsevier, 1967, pp. 267–343.

[218] F Zwicky et al. "Catalogues of Galaxies and Clusters of Galaxies (California Institute of Technology, Pasadena)". In: *Figure Captions* (1961).

Appendix A

Statistical Methods

A.1 Wiener Filter Review

Wiener filtering (WF) [211, 164] is a popular way to non-parametrically reconstruct cosmological data as, in the linear case, it should minimize variance. In the absence of non-Gaussian sources of signal or noise, WF is optimal in the sense that it is equal to the maximum posterior probability estimator [174]. Here we want to reconstruct the field itself, $\boldsymbol{s}$, given the noisy and/or poorly sampled data, $\boldsymbol{d}$. We define our estimated field $\hat{\boldsymbol{s}} = \boldsymbol{\Phi d}$, where $\boldsymbol{\Phi}$ is a linear operator, i.e. a N x M dimensional matrix transforming from "image space" to "field space". This can be found by attempting to minimize the variance of the residual

$$\langle (\boldsymbol{s} - \hat{\boldsymbol{s}})(\boldsymbol{s} - \hat{\boldsymbol{s}})^\dagger \rangle \tag{A.1}$$

with respect to $\boldsymbol{\Phi}$. The Wiener filtered estimator is

$$\hat{\boldsymbol{s}} = \boldsymbol{\Phi d} = \langle \boldsymbol{s d}^\dagger \rangle \langle \boldsymbol{d d}^\dagger \rangle^{-1} \boldsymbol{d} = \boldsymbol{S R}^\dagger \boldsymbol{C}^{-1} \boldsymbol{d}, \tag{A.2}$$

which will result in a variance of residuals of the form

$$\langle (\boldsymbol{s} - \hat{\boldsymbol{s}})(\boldsymbol{s} - \hat{\boldsymbol{s}})^\dagger \rangle = \boldsymbol{S} - \boldsymbol{S R}^\dagger \boldsymbol{C}^{-1} \boldsymbol{R S}. \tag{A.3}$$

Wiener filter only uses the mean and variance of the statistical distribution. If our underlying field is strongly non-Gaussian the WF may no longer be optimal. However, it will still minimize the variance, as defined in equation 4.6, just that this minimization of variance may not correspond to the notion of the best reconstruction as it only captures the two point statistics of the underlying field. It may be difficult to even define a measure to use for optimally of reconstruction in these cases, although nonlinear reconstruction methods certainly exist [174].

We now want to connect the Wiener Filter solution to the optimal power spectrum estimator. We multiply equation 4.7 by the Hessian matrix,

$$\begin{aligned} (\boldsymbol{F}\Theta)_l &= \frac{\boldsymbol{F}}{2} \textstyle\sum_{l'} F^{-1}_{ll'} (\boldsymbol{d}^\dagger \boldsymbol{C}^{-1} \boldsymbol{Q}_{l'} \boldsymbol{C}^{-1} \boldsymbol{d} - b_{l'}), \\ &= \frac{\delta_{ll'}}{2} (\boldsymbol{d}^\dagger \boldsymbol{C}^{-1} \boldsymbol{\Pi}_{l'} \boldsymbol{R R}^\dagger \boldsymbol{\Pi}_{l'} \boldsymbol{C}^{-1} \boldsymbol{d} - b_{l'}). \end{aligned} \tag{A.4}$$

The noise $\boldsymbol{b}$ can be similarly transformed as

$$b_l = \mathrm{tr}(\boldsymbol{\Pi}_l \boldsymbol{R}^\dagger \boldsymbol{C}^{-1} (\boldsymbol{N}) \boldsymbol{C}^{-1} \boldsymbol{R} \boldsymbol{\Pi}_l), \tag{A.5}$$

and the Hessian matrix itself as

$$F_{ll'} = \mathrm{tr}(\boldsymbol{C}^{-1} \boldsymbol{Q}_l \boldsymbol{C}^{-1} \boldsymbol{Q}_{l'}) = \frac{1}{2} |\boldsymbol{\Pi}_l \boldsymbol{R}^\dagger \boldsymbol{C}^{-1} \boldsymbol{R} \boldsymbol{\Pi}_l|^2. \tag{A.6}$$

Both the Wiener Filter and the optimal power spectrum estimator first weigh the data by the inverse covariance matrix, essentially down weighting modes that either have high measurement error or strong correlation with other measurements.